＜ 修 士 論 文 ＞

コールセンターオペレーターの音声品質評価の自動化の研究

滋 賀 大 学 大 学 院
デ ー タ サ イ エ ン ス 研 究 科
デ ー タ サ イ エ ン ス 専 攻

修了年度：２０２１年度

学籍番号：6020108
氏　　名：柴田　忠彦
指導教員：市川 治

提出年月日：２０２２年1月１２日

目次

第 1 章 はじめに

1.1 研究の背景

消費者の価値観が多様化する中、顧客が企業との多様な接点や商品・サービスを通じて得られる「体験価値」を高めることは、企業において重要な差別化要素となっている。E-mailやチャットボットなどを通じた顧客応対サービスが増える一方で、コールセンターは顧客ニーズを直接ヒアリングできる重要な場であり、顧客にとっては企業のブランドイメージを決定づける場となっている。

パソナグループのビーウィズ株式会社（本社：東京都新宿区、代表取締役社長：森本 宏一）は、2019年11月27日（水）に滋賀大学（滋賀県彦根市、学長：位田 隆一）と連携協定を締結した。ビーウィズが運営するコールセンターの応対品質の向上を目指し、コールセンター応対におけるオペレーターの「声の印象評価システム」について、データ解析を活用した研究を開始した。

ビーウィズは、2000年の創業よりコールセンターのアウトソーサーとして、企業のお客様対応窓口などのコールセンター運営業務を行っている。コールセンターでは、従来からオペレーターの応対音声を統一基準で評価する「モニタリング」を行うことで品質を維持・改善している。この「モニタリング」の効率化と精度向上を目指し、ビーウィズは2019年6月からAIによるコールセンター応対音声のリアルタイムテキスト化を活用した「応対評価の全件自動化」を開始している。これにより、オペレーターの「特定ワードの発話タイミングや回数」、「文字量」や「速度」等の11項目の自動評価を実現する一方で、「発声や発音」「声の表情」など、テキストでは表現されない応対評価項目はまだ自動化されていない状況である。

この連携協定締結により、ビーウィズがコールセンターの現場で積み上げた知見と、滋賀大学データサイエンス学部の科学的アプローチを組み合わせ、これまで人が評価してきたコールセンターにおけるオペレーターの「声から感じる応対の印象」を科学的に解析することで、応対評価と教育のサイクルを高速化し、コールセンターサービスの応対品質の向上を目指している。

当論文は、ビーウィズ株式会社との共同研究において「声から感じる応対の印象評価」を自動化することを目的にその要素技術について論ずる。

■ ビーウィズ株式会社 概要

2000年にコンタクトセンター、BPO(ビジネス・プロセス・アウトソーシング)エージェンシーとして創業。業界問わず、大手企業のコンタクトセンターBPOセンターを受託し、企業の顧客接点の最前線でサービス提供を行っている。近年は、グループ企業のアイブリット社で開発しているコールセンターシステム「Omnia　LINK（オムニアリンク）」を活用し、リアルタイムテキスト化機能を活用したFAQリコメンデーションソリューション「seekassist（シークアシスト）」やお客様の声の解析「VoCアナリティクス」など、新たなコミュニケーションデザインの提供を行っており、企業の競争力強化という切り口から、コンサルティ

ングからオペレーションまで幅広い領域で企業支援を行っている。

コールセンターの応対品質とは電話対応の適切さのことである。もう少し詳しく説明すると、オペレーターが電話対応をするときには基本的な知識や正確性などの業務遂行力と言葉遣いや傾聴力、共感力といったソフト面のどちらも持ち合わせている必要がある。応対品質が低いと顧客が不満を抱える可能性があり、顧客満足度の低下やブランドイメージのダウンに繋がる。逆に言うと、応対品質を高い基準でキープできていれば、顧客満足度アップや円滑なコールセンター運用ができるようになる。応対品質の確認方法としては、オペレーターが電話対応に必要な業務遂行力と言葉遣いや共感力、提案力などのソフト面を総合的に把握し、顧客が快適に使える品質基準に達しているのか確認する。応対品質を調査する方法には、下記のような3つの調査方法がある。

- モニタリング
  オペレーターの電話対応をチェックして、回答のスキルや顧客への話し方、マナーを調査する方法
- ミステリーコール
  調査業者から依頼を受けた第三者が顧客のふりをしてオペレーターの応対品質を調査する方法
- アンケートサービス
  オペレーターとの電話が終わった後に顧客に対してアンケート回答を依頼する方法

上記の方法にて、「応対の受け答えは迅速か」「顧客のニーズに合う情報を提供できているか」「顧客に寄り添うフレーズを利用できているか」など、コールセンターごとに重要視しているチェック項目を設けて評価していく。応対品質が悪いと、

- コールセンターに電話をしても目的達成や課題解決ができない
- オペレーターの対応が雑で不満がある
- オペレーターの言葉遣いや不機嫌な雰囲気が気になる

など顧客の不満や苦情の原因となり、顧客満足度の低下に繋がる。逆に応対品質が高いと、顧客に安心して問い合わせできる環境を作ることができ、顧客満足度の向上やブランドイメージアップに繋げることができる。コールセンターを運営していくうえで応対品質の向上は重要な課題となるため、自社の応対品質を正確に把握して改善していくことが求められる。

そのような環境の中で，コールセンターを運営する企業では顧客対応の中心となる優秀なオペレーター（顧客対応を担当する要員）の確保と育成，各種業務の効率化が喫緊の課題となっている。それらの課題に対して IT 技術の活用は有効なソリューションとなりうる。例えば応対品質評価の自動化である。応対品質に対して現在は主に各コールセンターの管理者（スーパーバイザーと呼ばれる，オペレーターの育成や管理を担うマネジメント職種）が人手で行っている。具体的にはコールセンターへの受電から通話終了までの数分から数十分に

わたるオペレーターと顧客との対話が録音された音声を管理者が耳で聞き，オペレーターの発話内容が聞き取りやすいか，声の大きさや抑揚の強さが適切かどうか，相手に良い印象を与える声の表情で話しているかなど数十項目の評価項目に対して手作業でスコアを付けている。管理者は公正かつ客観的な評価を行うために専門の教育を受け，評価基準のブレを防ぐために定期的な校正（キャリブレーション）も受けている。この評価作業をコンピューターによって自動化すれば，人手では困難な，あるいは実現不可能な以下の事柄を実現できる。

① 全数評価・・・ひとつのコールセンターにおける１ヶ月あたりの応対通話時間は合計で数千時間から数万時間に及び，その全てに対して人手で評価を行う事は不可能である。そのため，人手による評価は全ての録音データの中から一部をピックアップして行われている。それに対して，十分な性能を持つコンピューターであれば全数評価が可能となる。全数評価を行う事により，たまたま応対が良かった／悪かったといった偶然性を排除して公平な評価ができる。

② 一定の基準による評価・・・定期的に校正（キャリブレーション）を受けたとしても，管理者は人間である以上，その日の体調等により評価にばらつきが出る可能性は否めない。それに対してコンピューターは常に一定の基準で評価を行う事ができる。

③ 人が行う作業の肩代わり・・・管理者による評価作業の一部または全部をコンピューターに肩代わりさせる事により，管理者は他の作業に時間を割く事ができる。もしくは今までよりも少ない人員で業務を遂行する事ができる。

これらを実現する事により，以下の効果が期待できる。

・オペレーター個々のスキルを正しく把握する事による適切な改善指導の実現

・客観的で公平な評価によるオペレーターのモチベーションの維持，向上

・評価を行う管理者の肉体的，精神的な負担の軽減

・管理者が評価作業に掛ける時間を減らし，その代わりに現場でのオペレーターへの指導や援助を充実させる事によるサービスの品質向上

・評価作業に掛かるコストの削減

上記の通り，コールセンターの応対品質評価の自動化にはメリットが多い。そのため，これまでにも自動化の試みは行われ，製品やソリューションとして販売されているケースも存在する。例えば富士通株式会社や株式会社日立情報通信エンジニアリングの例では，発話速度，発話かぶりの回数や時間，必須ワードや NG ワードの使用回数を通話録音データから抽出し，コンピューターによる自動評価を行っている[1][2]。しかし，これらは主として音声データから音声認識技術によってテキスト情報を抽出し，音声とテキストのアラインメントを取る（対応づける）事によって実現していると考えられ，先に述べたような「相手に良い印象を与える声の表情で話しているか」といった，言語解析型の音声認識技術だけでは実現できない音響的な評価指標に対する自動評価が行われているものは現時点では見当たらない。

1.2 研究の目的

本研究では，ビーウィズ株式会社が行っているコールセンターにおけるオペレーターの応対品質評価のうち，自動化が行われていない評価項目の自動化に向け， その実現可能性の検討を行う。評価項目は表１の通りである。本研究では評価項目１から１８の自動化に向けて研究を行った。実現手段としては音響解析型の技術を応用し，人手による評価と比較してどの程度の精度（再現率，適合率）で自動評価ができるかを確かめる。その結果が人手による評価よりも劣る場合は，自動評価の実現に向けてどのような課題があり，どのような解決手段が考えられるかを検討する。なお、評価項目20「全体として表情があり、感じがよいか」については、滋賀大学データサイエンス研究科修士課程の令和2年卒である高山が音声感情認識技術を活用した研究をしており、データの前処理や音響特徴量の抽出方法等において、本論文ではその成果[29]を参考として研究を行っている。

表1 コールセンターの応対品質評価の項目

No.	分類	評価項目
1	大きさ	全体的に大きすぎないか
2		全体的に小さすぎないか
3		音量の変化が不自然ではないか
4	語頭	語頭が弱くないか
5	語尾	語尾の跳ね
6		語尾消え（語尾の明瞭さ）
7		語尾伸び
8		語尾上がり
9		語尾下がり
10		語尾の強さ
11	滑舌	全体的に滑舌は悪くないか
12		特定の箇所のみ滑舌が悪い
13	抑揚	抑揚が極端でないか
14		抑揚が弱すぎないか
15		抑揚の場面が適切か
16	スピード	全体のスピードの速さ
17		全体のスピードの遅さ
18		スピードの変化
19	表情	一部表情がミスマッチな箇所が無いか
20		全体として表情があり，感じが良いか

1.3 本論文の構成

本論文は全 5 章からなる。その構成は以下のようになっている。第 1 章では，本研究の背景としてコールセンターにおけるオペレーターの応対品質評価の現状と課題，自動化によって期待される効果をまとめ，本研究の目的を述べた。第 2 章では，応対品質評価の自動化においては音響解析型のアプローチで行う為、音声から得られるパラ言語情報と音響特徴量の概要を述べ、また、実験環境についてツールを使用しているためその概要を述べる。第 3 章では，音響特徴量を使用した提案法によって、コールセンターオペレーターの印象評価に対して推定した実験の内容及びその結果を示す。第4章では、本研究全般で課題であるデータ不均衡問題における対策及びその実験の結果を示す。第 5 章では、本研究のまとめと考察，今後の課題について述べる。

第２章 音響特徴量の概要と実験環境について

2.1 パラ言語について

この節では、自動化対象である評価項目1〜18（声の大きさ、語尾、抑揚、滑舌、話速）と重要な関係のあるパラ言語についてその概要を述べる。

パラ言語とは、コミュニケーションの際に言語情報を補う言語以外の音声のことで、簡単に言うと、話す速さ、声の強さ、高さ、沈黙、イントネーションなどのことである。そもそも人のコミュニケーションは、言語によるものと、それ以外によるものに二分できるが、この言語以外によるコミュニケーションのことを非言語行動（ノンバーバル・コミュニケーション）と言う。非言語行動は、コミュニケーションで伝えられる情報の70%を占めるとする研究もあり、円滑なコミュニケーションするうえで大変重要な要素である。非言語行動には、ジェスチャーや目線、表情、うなずきなどの身体動作に関するものと、相手との距離に関するもの、そして音声に関するものがある。この音声に関するものをまとめてパラ言語と言い、言語情報を補う音声ということから周辺言語とも言われている。

パラ言語は、コミュニケーションで使われる言語以外の音声、と書いたが、具体的にどのようなものがパラ言語にあたるのであろうか。主に以下に挙げるものが、パラ言語に含まれるとされている。

話す速さ、声の高さ、強さ、声色、イントネーション、咳払い、ため息、笑い、相槌、フィラー（「えー」、「あのー」など）、沈黙　等

沈黙は、何もしていないからパラ言語ではない、と思われるかもしれないが、沈黙によって相手への反発を示す、相手の発言を反芻する、自分の発言を準備するといった意味があるため、パラ言語に含まれる。パラ言語は年齢や性別などによって個人差があり、また同じ人でも話す内容や相手、場面によって変化する。その時の感情も反映されるので、例えば話している間に不安になれば、沈黙や話し方の乱れが増える、というようなことが起こる。

このように、話し手の感情や意図によって意識的に使われる言語以外の表現がパラ言語にあたる。

一方で、言語の意味に関わる以下のようなものはパラ言語に含まれない。

・アクセント

アクセントは言語によって決められているものなので、パラ言語には含まれない

例えば「雨（あめ）」と「飴（あめ）」では、音の高さ（高低アクセント）によって意味の違いがあり、これは相手や場面によって変わるものではないため、パラ言語とは言えない。

・相槌の一部

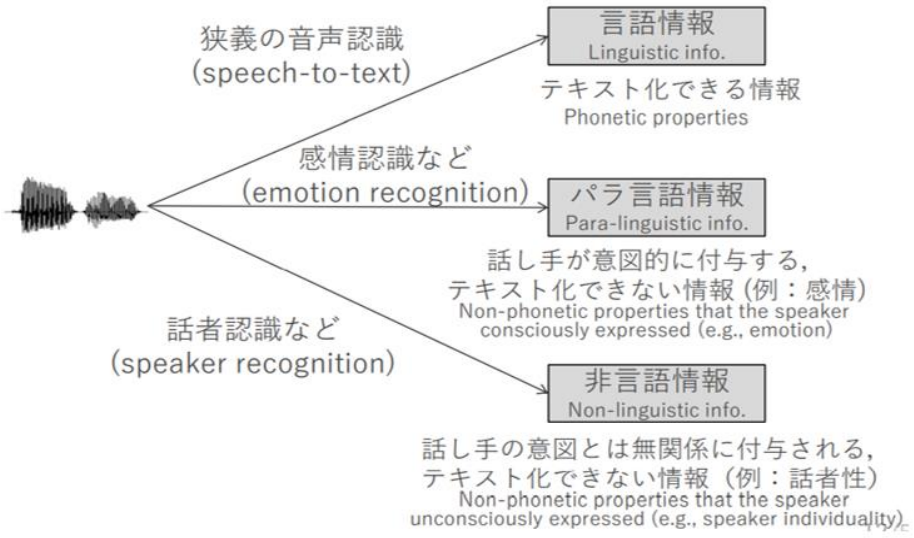
また、「へえ」や「そう」などの言語的意味を持つ一部の相槌もパラ言語に含まれない。例えば日本語の相槌「へえ」には、「知らなかった！」「初めて聞いた！」というような

意味がある。このような意味を持つ相槌はパラ言語ではなく、言語であるとされる。同じ相槌でも、「ええ」とか「あぁ」などはパラ言語である。

先ほど、パラ言語は年齢や性別などによって個人差があると書いたが、文化によっても感じ方、捉え方に違いがある。たとえば、一般的に日本人の英語の話し方は単調に聞こえる場合があり、アメリカ人にとっては冷たいとか興味がないと感じられることがある。また、日本語学習者でも疑問文のイントネーションが違っている（文末が下降イントネーションになる）ために、相手を責めているように聞こえてしまうことがある。このように、パラ言語によって相手に悪い印象を与えてしまったり、勘違いされてしまったりする場合がある。

図１のように、音声データから得られる情報は大きく3つに別れ、本節でここまで述べた通り、今回の印象評価の研究においては、言語情報だけでなくパラ言語情報が相手に与える印象がとても重要である。

図1　音声の持つ情報



2.2　音響特徴量

この節では本研究で使用する音響特徴量について述べる。特にopenSMILE ツールキット（音声信号から特徴量を抽出できるオープンソースのツールキット）と呼ばれるツールにより取得できるIS09,IS10特徴量を利用するためそれについて述べる。

音声からテキスト情報を抽出する通常の音声認識では，音声ファイルから波形データが窓掛け操作によりフレーム毎に切り出され，離散フーリエ変換によって周波数スペクトルに変換される。その後， 周波数毎のパワーに変換され，微細構造を落とすメルフィルタバンクを掛ける事で得られる対数メルスペクトルや，更に離散コサイン変換を行う事で得られる MFCC が音響特徴量として標準的に用いられる[15]。これらは人の声道特性（≒言葉を話している人間の喉頭から口蓋，鼻腔の形）を良く表す特徴量である。音響解析型で用い

られる音響特徴量は，この対数メルスペクトルや MFCC に加えて，声の高さを表す基本周波数，音量，音声波形の揺らぎを表すシマーやジッタなど，様々な特徴量が使用される。感情等の人のパラ言語情報は声道の形だけに表れるのではないと考えられる為であり，国内外の多数の研究者により多様な特徴量が提案されている[9][16][17][18][19]。これらのフレーム毎に抽出される特徴量を LLD（LowLevel Descriptor）と呼ぶ。

このように、パラ言語情報で用いられる音響特徴量は多数の研究者により多様な特徴量が提案されているが，その中でも音声感情認識等でよく用いられている特徴量セットが INTERSPEECH 2009 Emotion Challenge（IS09）特徴量セット[16]と INTERSPEECH 2010 Paralinguistic Challenge（IS10）特徴量セット[17]である。これらは音声言語処理分野で一流の国際会議である INTERSPEECH で提案された特徴量セットで，性能が良い上に openSMILE ツールキット（音声信号から特徴量を抽出できるオープンソースのツールキット）によって簡便に特徴量を抽出できる事から広く用いられている。この特徴量では各 LLD に対して発話全体の平均や分散，線形回帰直線の傾きや切片などの統計量を計算したものを機械学習モデルの入力として用いる。これは，音声に含まれる感情は音声波形の 1 フレーム（10～数十マイクロ秒）や数フレーム程度の瞬間的な値では捉えることができず，発話全体の傾向を見ることによって捉えることができると考えられている為である[9][20]。複数の LLD のそれぞれに対して複数の統計量を計算して使用するため，音声感情で用いられる特徴量は数百～数千次元の特徴量ベクトルとなる事が多い。音声感情認識では明確に必要な特徴量が未だ判明していないため，関係があると想定される特徴量を全て使用する事が一般的に行われている[20]。

表2 に IS09 特徴量セットの内容を，表3 に IS10 特徴量セットの内容を記載する。本研究ではこれら 2 種類のセットに含まれる一部の特徴量を利用しモデルを構築している。具体的にどの特徴量を各評価項目の自動推定で使用したかは第3章の各評価項目に対する自動推定の実験にて述べる。

表 2　INTERSPEECH 2009 Emotion Challenge（IS09）特徴量セット

LLD	**音量**，F0，MFCC（1-12 **次**）， **その時点での音が声である確率**， **波形のゼロ交差率** **※上記の静的特徴量および Δ（1 階差分）**
統計量	**算術平均**，**標準偏差**，**尖度**，**歪度**， **最大値**，**最小値**，**最大値と最小値の差分**， **最大値位置**，**最小値位置**， **線形回帰直線の傾きと切片**， **線形回帰直線からの二乗誤差**，
次元数	384 **次元**

表 3　INTERSPEECH 2010 Paralinguistic Challenge（IS10）特徴量セット

LLD	メル周波数帯（0-7 次）の対数パワー，MFCC（0-14 次）， LSP（線スペクトル対）周波数（0-7 次）， F0，F0env（F0 の包絡），ラウドネス， シマー（振幅方向の波形の揺らぎ）， ジッタ（時間軸方向の波形の揺らぎ）， 差動フレーム間ジッタ（ジッタのジッタ）， 有声音らしさ ※上記の静的特徴量および Δ（1 階差分）
統計量	算術平均，標準偏差，尖度，歪度， 四分位数，四分位間の範囲， 最大値位置，最小値位置， 線形回帰直線の傾きと切片， 線形回帰直線からの線形誤差と二乗誤差， 1%，99%パーセンタイル， 99%パーセンタイルと 1%パーセンタイルの幅， レンジの 75%を超えている時間の割合， レンジの 90%を超えている時間の割合， 入力の総継続時間 F0 のオンセット数（疑似的な音節数）
次元数	1,582 次元

2.3 Wekaについて

本研究では、データマイニングツールWekaを使用しているため、その概要を述べる。WEKA は、ニュージーランドのワイカト大学（University of Waikato）のIanh. Witten、Eibe. Frank を中心とした機械学習の研究者によって開発され続けている、Java 言語によるオープンソースのデータマイニングのフリーソフトである。実は、WEKAはニュージーランドに生息し、飛ぶのが苦手であるが、探究心が非常に強い鳥の名前である。ソフトWEKAに関する1次情報は、次のページから入手できる。

（http://www.cs.waikato.ac.nz/~ml/WEKA/）

上記のサイト、あるいは次のサイトからコンピューターのOS(Windows、Mac、Linuxなど)にマッチしたWEKAを入手することができる。

（http://prdownloads.sourceforge.net/WEKA/WEKA.ppt）

使用しているマシンに Java 言語がインストールされていない場合は、Java が同梱されているバージョンを選んだ方がよい。WEKAで扱っているデータマイニングのアルゴリズ

ムの基礎に関する本としては、ソフトの開発者の著書（参考文献[21]）がある。WEKAは、データの前処理、分類と予測、クラスタリング、相関ルール、視覚化に関するアルゴリズムの集合体である。WEKAでは、データセットの中の列(変数)を属性（attribution)、行(個体)をインスタンス（instance)、特定のタスクを実行するアルゴリズムの集まりをスキーム(scheme)、判別・分類を行うスキームを分類器（classifier)、樹木モデルを決定木(decision tree) と呼ぶ。本稿で用いたWEKAは Windows 用のバージョン3-8-1である。

WEKAのダウンロードとインストールの手順は紙面の都合により割愛する。WEKAを起動すると図２のようなGUI画面が開かれる。GUIの鳥がWEKAである。GUIの右側には5つのボタンがある。それぞれのボタンを押すとデータを操作するパネルが開かれる。その主な機能を表４に示す。

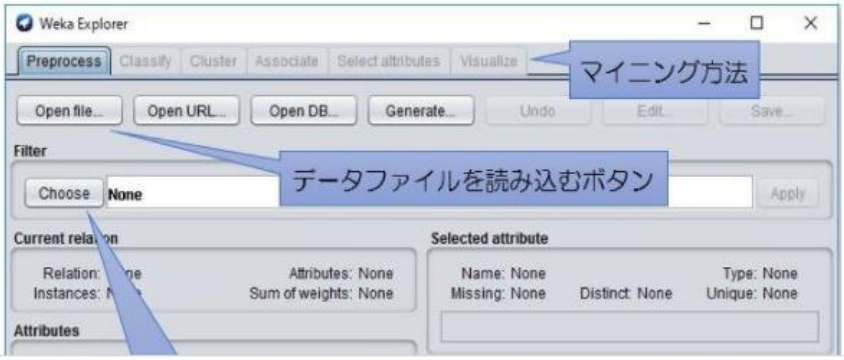
図２　wekaのGUI画面

表4　GUIのボタン機能

Explorer	メニュー選択型の操作環境
Experimenter	学習スキームの間の統計検定などを行う環境
KnowledgeFlow	データ処理・マイニングのプロセスをアイコンで連結してマイニングを行うGUI環境
Workbench	ワークベンチ
SimpleCLI	コマンドによる操作環境

WEKAを扱っている書籍 “Data Mining”（参考文献[21]）ではコマンドラインを用いて解説しているが、本稿ではGUIの Explorer 環境を用いることにする。WEKAのGUIにおける [Explorer] ボタンを押すと図3のパネルが開かれる。

図３　Explorer 画面

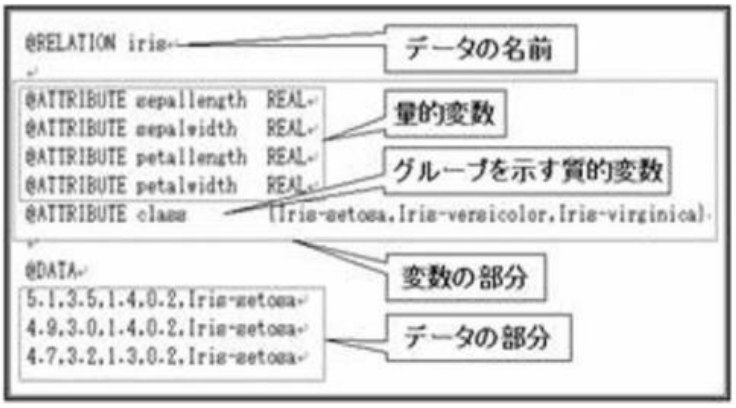


Explorerのパネルの上部には6つのタブが設置されている。WEKAのデータ処理・マイニングに関する機能はこの6つのタブに分類されている。この6つのタブに含まれている機能及び主なスキームを表5に示す。WEKAは、カンマ区切りのCSV形式、C4.5形式、ARFF形式などを読み込むことができる。　CSV形式は表計算ソフトExcelでも簡単に作成できる。ARFF形式のデータファイル概観を示すため、データセットIris のARFF形式の一部分のコピーを図4に例示する。ARFFファイルは、データの属性部とデータの部分に分けて記述する。属性の部分は、属性のデータの性質について具体的に記述し、データの部分は属性の順序の順にカンマでデータを区切る。データはローカルディスク、インターネット、データベースから直接読み込むことができる。

表5　GUIのボタン機能

Preprocess	データの選択と修正などのための前処理に関するフィルタ環境で、 44種類(supervised 7種類、unsupervised 37種類)のアルゴリズムがある
Classify	分類と予測に関する環境で、71種類(Bayes 7種類、function 12種類、lazy 5種類、 meta 23種類、misc　3種類、trees　11種類、rules 10種類)のアルゴリズムがある
Cluster	クラスタに関する環境で5種類のアルゴリズムがある
Associate	相関ルールに関する環境で、3種類のアルゴリズムがある
Select attributes	属性の選択に関する環境で、20種類(Attribute Evaluator 12種類、 search Method 8種類)のアルゴリズムがある。
Visualize	データの2次元グラフの環境である

図4　ARFFファイルの形式



2.5 本章のまとめ

本章では自動化対象である評価項目1～18（声の大きさ、語尾、抑揚、滑舌、話速）と重要な関係のあるパラ言語についてその概要を述べた。また、本研究で使用する音響特徴量について述べた。特にopenSMILE ツールキット（音声信号から特徴量を抽出できるオープンソースのツールキット）と呼ばれるツールにより取得できるIS09,IS10特徴量を利用するためそれについて述べた。最後に、データマイニングツールWekaを使用しているため、その概要を述べた。後続の第3章では、これら音響特徴量や実験環境を用いて印象評価の自動推定実験をおこなっている。

第３章　コールセンター音声の分析

3.1　　コールセンターの応対品質評価について

本研究の主題であるコールセンターにおけるオペレーターの応対品質評価について，現在人手で行われている評価の内容を説明し，本研究で何を目指すかを述べる。本研究で使用した音声データの提供元であるビーウィズ社では，「声の印象」に関して表 6 に示す20項目の評価を実施している。評価者はオペレーターと通話相手とのやり取りが録音された音声を耳で聞き，これらの評価項目一つひとつに対して以下の 3 段階の評点を付与している。

・評点「1」：相手の心情を害するおそれがある

・評点「2」：改善ポイントあり

・評点「3」：適切な応対範囲

表 6　　コールセンターの応対品質評価の項目

No.	分類	評価項目
1	大きさ	全体的に大きすぎないか
2		全体的に小さすぎないか
3		音量の変化が不自然ではないか
4	語頭	語頭が弱くないか
5	語尾	語尾の跳ね
6		語尾消え（語尾の明瞭さ）
7		語尾伸び
8		語尾上がり
9		語尾下がり
10		語尾の強さ
11	滑舌	全体的に滑舌は悪くないか
12		特定の箇所のみ滑舌が悪い
13	抑揚	抑揚が極端でないか
14		抑揚が弱すぎないか
15		抑揚の場面が適切か
16	スピード	全体のスピードの速さ
17		全体のスピードの遅さ
18		スピードの変化
19	表情	一部表情がミスマッチな箇所が無いか
20		全体として表情があり，感じが良いか

これらの評点を項目毎に重みづけして合計した点数が，音声の音響的な特徴から相手が受ける印象を表す「声の印象」の評価点となる。更に，対話の内容（オープニングの名乗りからクロージングまで適切な対応が出来ているか），オペレーターとして求められている問題解決力を発揮できているかなどの複数の評価項目について評点が付与され，オペレーターの総合的な評価点が決定される。

本章では，表 6 の No.1から18の評価の自動化を目指す。音響特徴量を使用して機械学習により評点を推定する分類器を作成し，人手による評価と比較してどの程度の精度（再現率，適合率）で推定が行われるかを確かめる。その結果が人手による評価よりも劣る場合は，自動評価の実現に向けてどのような課題があり，どのような解決手段が考えられるかを検討する。

3.2 使用したデータ

3.2.1 受領データ

ビーウィズ社から受領したデータについて説明する。以下の 2 種類のデータを受領した。

① コールセンターの応対を，受電から通話終了まで録音した wav 形式の音声ファイル。8kHz のサンプリングレートで収録されている。（受託業務の音声は含まない。）

② Excel 形式のラベルデータ。以下の情報が記載されている。

・発話の開始時刻と終了時刻．単位はミリ秒

・発話内容（音声認識により自動書き起こしされたテキスト情報）

・文意情報（文意情報は3.3.2.1節にて詳細に説明する）

・話者がオペレーターと通話相手のどちらであるかを表すフラグ

・20項目の評価項目毎に人手で付与された評点

尚，オペレーターと通話相手の音声が重なっている発話区間については「評価対象外」のラベルが付与されている。これは音声が重なっている場合に機械学習による推定が困難になると想定されたためビーウィズ社にて付与して頂いた。音声ファイルとラベルデータからは予め人名，住所，電話番号等の個人情報を削除された状態でデータの提供を受けた。ラベルデータ（Excelファイル）のイメージを図5に示す。

図 5　ラベルデータ（Excel ファイル）のイメージ

						1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
		総合			カテゴリ		大きさ			語調			語尾				滑舌			抑揚	
		72			評価項目	全体的に大きすぎないか	全体的に小さすぎないか	音量の変化が不自然ではないか	語調が強くないか	語尾の跳ね	語尾消え（語尾の明瞭さ）	語尾伸び	語尾上がり	語尾下がり	語尾の強さ	全体的に滑舌は悪くないか	し特定の箇所のみ滑舌が悪い	抑揚が極端でないか	抑揚が強すぎないか	抑揚の頻度が適切か
					通話計	87	87	86	85	86	87	82	87	86	87	83	81	87	86	76
					平均	3.0	3.0	3.0	2.9	3.0	3.0	2.8	3.0	3.0	3.0	2.9	2.8	3.0	3.0	2.6
開始時間	終了時間	お客様の		RecognitionText	合計	全体的(	全体的(	音量の変	語調が強	語尾のば	語尾消え	語尾伸び	語尾上が	語尾下が	語尾の強	全体的に滑	し特定の箇	抑揚が極	抑揚が強	抑揚の頻
0:00:00.906	0:00:03.367		1	お電話ありがとうございます無事でございます	59	3	3	3	3	3	3	3	3	3	3	3	2	3	3	3
0:00:04.618	0:00:10.845		2	[redacted]																
0:00:11.105	0:00:12.051		1	かしこまりました	58	3	3	3	3	3	3	3	3	3	3	3	2	3	3	2
0:00:12.355	0:00:17.662		1	今確認いたしますので恐れ入りますがお名前をフルネームでお	57	3	3	3	2	3	3	2	3	3	3	3	2	3	3	3
0:00:18.447	0:00:18.947		2	[redacted]																
0:00:19.342	0:00:20.297		1	かしこまりました	56	3	3	3	3	3	3	3	3	3	3	2	3	3	3	2
0:00:20.617	0:00:23.670		1	ご応募いただいた際のお電話番号よろしいでしょうか	57	3	3	3	3	3	3	2	3	3	3	2	2	3	3	3
0:00:25.414	0:00:27.550		2	[redacted]																
0:00:27.880	0:00:28.373		1	あっはい	60	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
0:00:29.247	0:00:29.375		2	[redacted]																
0:00:29.694	0:00:29.911		1	はい	60	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
			2	[redacted]																
0:00:30.893	0:00:31.297		1	はい	60	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
0:00:31.363	0:00:31.482		2	[redacted]																
0:00:31.783	0:00:33.499		1	はい、【フルネーム】様でいらっしゃいま	60	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
0:00:33.499	0:00:33.681	番号のあり	1	[redacted]																
0:00:33.499	0:00:33.681		2	[redacted]																

3.2.2 データの前処理

受領した音声データは受電から通話終了までが 1 つの wav ファイルに記録されていた。これを発話区間ごとに 1 つの wav ファイルとするため，ラベルデータに記載された発話区間の開始時刻と終了時刻の情報を元に音声加工・変換ソフトの sox を用いて wav ファイルの分割を行った。

3.3 提案法：応対品質評価の自動推定

ここでは、コールセンターオペレーターの声の大きさ、語頭、語尾、滑舌、抑揚、スピードに対する評価の自動推定方法と実験結果について記載する。（評価項目については表6を参照。

音量、語尾、滑舌、抑揚、話速の自動推定については、各評価項目に対する音響特徴量を作成し、SVM等の機械学習で分類問題として評点１， ２， ３を推定するモデルとして実装する。各評価項目に対する特徴量としては、音量であれば音圧（ボリューム）、語尾の上がり下がり等であれば発話末尾0.5秒の音量やピッチの変化量、滑舌はMFCC分散値、抑揚はピッチの分散、話速であれば単位時間当たりのモーラ数を考える。応対品質評価の自動推定のモデル概要が図6である。各評価項目に対する詳細内容は後述する。

図 6　音量、語尾、滑舌、抑揚、話速のモデル概要

音声から 音響特徴量 を抽出 ⇒ 教師あり機械学習

※各音響特徴量
①音量：個別VOL、個別VOL/全体VOL平均、個別VOL/お客様全体VOL平均
②語尾：発話末尾0.5秒の音量変化量、発話末尾0.5秒のピッチ変化量　等
③滑舌：mfcc平均、mfcc分散、mfccΔ分散
④抑揚：ピッチ最大値、ピッチ最小値、ピッチΔ分散
⑤話速：発話時間、モーラ数、モーラ数/発話時間

3.3.1 音量の自動推定

ここでは、音量の自動推定について記載する。声の大きさは、音響特徴量としては音圧として現れる。音圧はIS09（opensmile特徴量）にRMSenagyという特徴量で含まれているため、まずは2.2節で述べたIS09（opensmile特徴量）にて音量の評価が自動推定できるか実施した。機械学習の手法として最初はSVM（サポートベクターマシン）を使用した。実験環境は2章で紹介したデータマイニングツールWekaを活用した。初期にSVMを採用

した理由としては、SVM では，特徴次元が増えても分類器が有効に働くことが知られている[22]。

表7は「音量が大きすぎないか」について評価者がラベリングした評点１（大きい）、評点２（やや大きい）、評点３（問題なし）に対して推定した結果である。検証方法はk-分割交差検証法（k-Fold-CV）で実施した。全体の正解率は94.1%とかなり高い精度となっているが、評点１（大きい）の再現率が低い結果となった。

表 7　「音量が大きすぎないか」SVMによる初期実験

SVM+opensmile

		推定結果		
		1	2	3
正解	1	2	6	3
	2	1	66	50
	3	0	18	1175

再現率	適合率	F 値
0.182	0.667	0.286
0.564	0.733	0.638
0.995	0.957	0.971

正解率
94.1

IS09（opensmile特徴量）は特徴量次元が多過ぎるため、音圧（RMSenagy）に関する特徴量に絞ることにした。また、評価者のコメントを確認すると「全体の音量に対してその発話の音量が大きいか」「お客様の声に対して発話の音量が大きいか」というのが評価に重要であることから、以下の特徴量で実験することにした。

① オペレーター個別発話音量/オペレーター全体発話平均音量
② オペレーター個別発話音量/カスタマー全体発話平均音量
③ オペレーター個別発話音量

上記①，②の特徴量が評点と相関があるかを確認してみたところ、図7,図8の通り評点が低くなるほど①，②の特徴量が大きくなり、相関があることがわかる。

図7　オペレーター個別発話音量/オペレーター全体発話平均音量

青：評点１　　赤：評点２　　水色：評点３

・評点が高いと分布が左より
・評点が低いと分布が右より
⇒顧客音量に対してオペレータ音量が大きいと評点が低くなる傾向

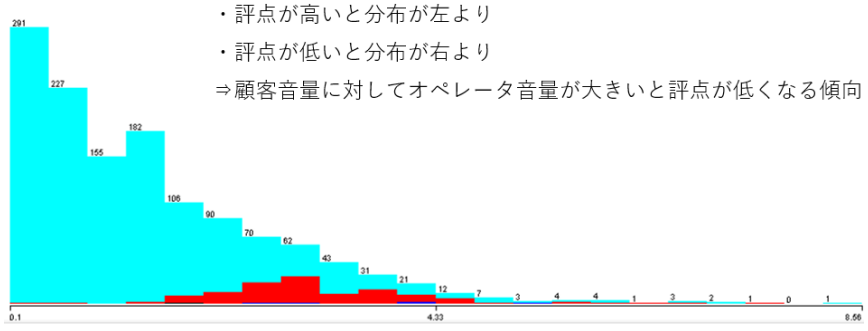
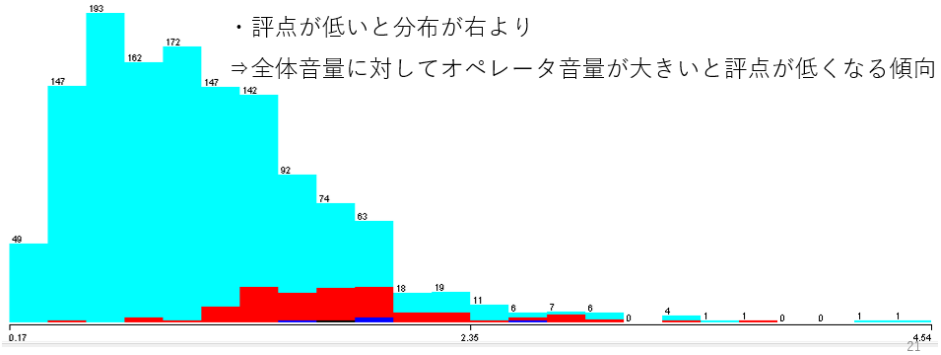


図8　オペレーター個別発話音量/カスタマー全体発話平均音量

青：評点１　　赤：評点２　　水色：評点３

・評点が高いと分布が左より
・評点が低いと分布が右より
⇒全体音量に対してオペレータ音量が大きいと評点が低くなる傾向



この３つの特徴量でSVMにて評点１，２，３の分類器を作成し実行したところ、IS09（opensmile）全特徴量で実施した結果より精度が下がってしまった。そこで手法を変え、決定木で再度実行したところ、IS09（opensmile特徴量）全特徴量で実施した結果に対して精度が上がることが確認できた。その実験結果が表8、表9である。但し、評点１の再現率が低い結果となった。これは、評点１の件数が少なくデータ不均衡による可能性がある。データ不均衡対策による精度の改善案については第4章で述べる。

表8　　「声が大きすぎないか」の実験比較結果

SVM+opensmile

		推定結果		
		1	2	3
正解	1	2	6	3
	2	1	66	50
	3	0	18	1175

再現率	適合率	F 値
0.182	0.667	0.286
0.564	0.733	0.638
0.995	0.957	0.971

正解率
94.1

SVM+3つの特徴量

		推定結果		
		1	2	3
正解	1	0	7	4
	2	2	21	94
	3	0	1	1192

再現率	適合率	F 値
0	0	0
0.179	0.724	0.288
0.999	0.924	0.893

正解率
91.82

決定木+３つの特徴量

		推定結果		
		1	2	3
正解	1	3	7	1
	2	2	92	23
	3	0	35	1158

再現率	適合率	F 値
0.273	0.6	0.375
0.786	0.687	0.733
0.971	0.98	0.975

正解率
94.85

表9　　「声が小さすぎないか」の実験比較結果

SVM+opensmile

		推定結果		
		1	2	3
正解	1	0	7	3
	2	0	70	152
	3	0	67	1016

再現率	適合率	F 値
0	0	0
315	0.486	0.383
0.938	0.868	0.902

正解率
82.5

SVM+3つの特徴量

		推定結果		
		1	2	3
正解	1	1	5	4
	2	2	93	125
	3	3	112	966

再現率	適合率	F 値
0.1	0.167	0.125
0.423	0.443	0.433
0.894	0.882	0.88

正解率
80.8

決定木+３つの特徴量

		推定結果		
		1	2	3
正解	1	1	7	2
	2	1	108	113
	3	0	45	1038

再現率	適合率	F 値
0.1	0.5	0.167
0.486	0.675	0.565
0.928	0.9	0.928

正解率
87.2

3.3.2 語尾系の自動推定

ここでは語尾の自動推定について記載する。語尾は聞き手の印象に非常に重要な要素である。まず、よく見かけるのが「聞き取りにくい語尾」である。聞き取れない理由はさまざまで、単純に声が小さくて聞こえなかったり、早口すぎて内容が聞き取れなくなる人もいる。また、「○○なんですけど……」と文章の途中で言葉が終わってしまい、語尾そのものがなくなることもある。話者が自信なさげに見えると「発話内容に自信がない」と思われてしまい、聞こえにくい言葉は、聞き手のストレスを生む。「え、今なんて言ったの？」と思いながら聞き続けたり、確認のために質問したりするのは、聞き手からすれば手間がかかる行動である。日本語は、語尾で肯定文か否定文かが決まる。例えば「今期の売り上げは予算達成しました／しませんでした」「今回の施策を採用します／しません」など。「したのか、しなかったのか」と恐ろしいことに逆の話になってしまう。聞き手に誤った認識を持たせないように、相手が聞き取れない話し方は、避けなければならない。

次によく聞く語尾は「雑な語尾」。語尾までしっかり聞こえるように声を出すのは大切である。しかしその音が強すぎると、聞き手にきつい印象や失礼な印象を与えることになる。代表的な例としては「語尾だけ音が高くなり、ボリュームがぐっと大きくなる」、「"○○なんですよねー""○○なんですけどー"など、語尾を不用意に伸ばすことが多い」、質疑応答など聞き手からの発信に対して「"○○っすか？"（での声が小さく"っ"に聞こえる）と反応する」などである。このような語尾を使う話者からは、勢いを感じる反面、粗雑な印象を受けやすいのである。「こんな雑な対応をする人は、仕事に対し

てもいい加減なのではないか」など、聞き手に不信感を持たれてしまっては損である。特に聞き手からの質問や反応に対して、雑に見える反応をするのはいただけない。聞き手は自分の存在を大事にしてもらっていないように感じ、話者への好感度が下がってしまうであろう。人が不快に感じる話し方の語尾には例えば以下がある。

・語尾が伸びる

女子高生の話し方を例えに出すことが多い、「〇〇でぇ?」「〇〇にぃ?」というように語尾を伸ばして話す癖である。必ず語尾の後ろに母音（「ぇ」「ぃ」）が入ってしまう。ビジネスシーンだと、頼りないとか幼いという印象を与えてしまう。どんなに内容がしっかりした発話でも、話し方で損をしてしまうので伸ばさないようにしっかりと発音する必要がある。

・語尾が強い

語尾が強くなる癖の人は印象がきつくなってしまうことがある。それ以外にも変なリズムがついて聞きづらくなったりする。「〇〇で、」「〇〇に、」「〇〇の、」と句読点の前の言葉を強く言い、それが繰り返されると、聞いている人が話の意味を理解しづらくなる。語尾をソフトに話すように気をつける必要がある。

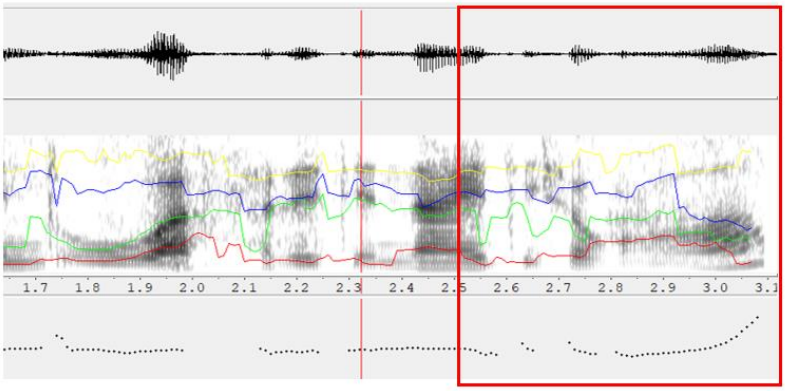
語尾の評価項目とそれに対する分析方針は表10の通りである。語尾跳ね、語尾消え、語尾伸び、語尾上がり、語尾下がり、語尾の強さという評価項目があり、例えばNo8語尾上がりでいえば、発話末尾のF0（ピッチ：声の高さ）がそれより前の発話時と比べ上昇しているかどうかが方針となる。基本的にはどの項目も発話末尾の音量またはF0（ピッチ）の変化をどうとらえるかが重要と考えた。

表10　　語尾系の分析方針

No	評価項目	分析方針
5	語尾跳ね	語尾のピッチが短時間で上昇
6	語尾消え	語尾の音量が発話全体と比較して小さい
7	語尾伸び	語尾のピッチが一定時間変化がない
8	語尾上がり	語尾のピッチが上昇
9	語尾下がり	語尾のピッチが下降
10	語尾強さ	語尾の音量が発話全体と比較して大きい

そこで、上記分析方針で問題ないかをF0や音量を可視化するツールを使用して確認してみた。使用したツールはWaveSurferである。WaveSurferはスウェーデンのKTHが開発・運用している音声分析ソフトウェアであり、図9のように上段に音声波形、中段に音声のスペクトログラム、下段にF0（ピッチ）の時間変化を表示してくれる。

図9　　WaveSurferの表示例



WaveSurferを使って、各評価項目に対する評点１のwaveファイルを入力に音声波形、スペクトログラム、F0の時間変化を表示してみた。それが図10から図15である。語尾跳ねでは、発話末尾の数フレームのF0が直前と比較して上昇しているのがわかる。語尾伸びでは、発話末尾のスペクトログラム（フォルマント）が一定時間継続している。語尾消えは発話末尾数フレームの音量が全体と比べ小さいのがわかる。語尾上がりは発話末尾の数フレームのF0が上昇しており、語尾下がりはその逆にF0が下降しているのがわかる。語尾の強さは発話末尾数フレームの音量が大きい。ここで、語尾跳ねと語尾上がりは共にF0の上昇という点で共通だが、語尾跳ねの方がより短時間での変化が激しいものと考える。このように、基本的には先で述べた方針通りの分析で問題ないことが確認できた。

図10　　WaveSurferの表示例（語尾跳ね評点１）

・評点１の例
・発話末尾の数フレームのピッチが直前と比較して上昇

発話内容：お試しセットのお届け日の事っていうことでよろしいですか

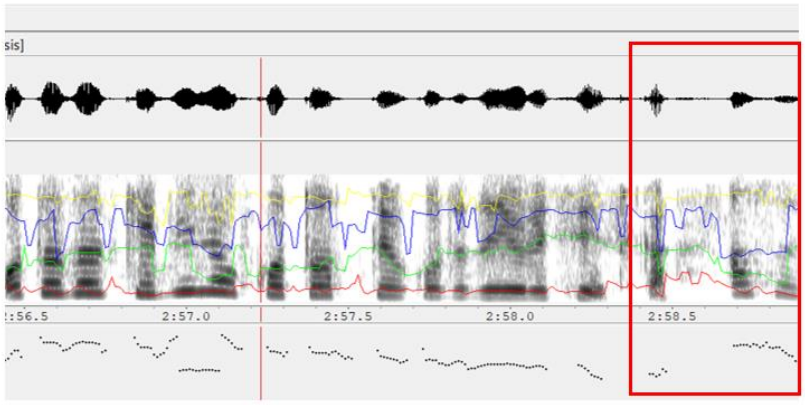


図11　　WaveSurferの表示例（語尾伸び評点１）

・評点１の例

・発話末尾のスペクトログラム（フォルマント）or音素が一定時間継続

発話内容：えーそうですねーーー

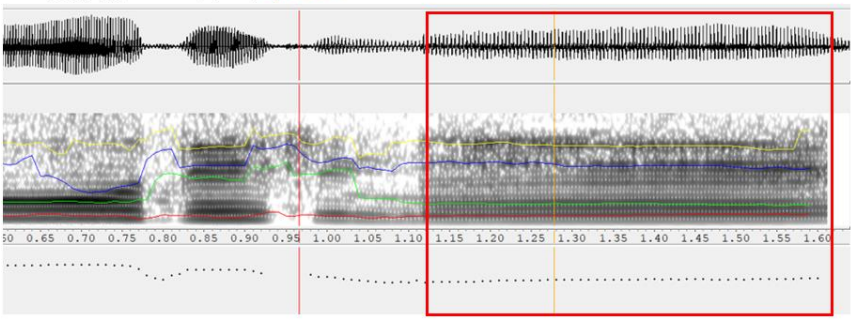


図12　　WaveSurferの表示例（語尾消え評点１）

・評点１の例

・発話末尾の数フレームの音量が全体と比べ小さい

発話内容：頂いております

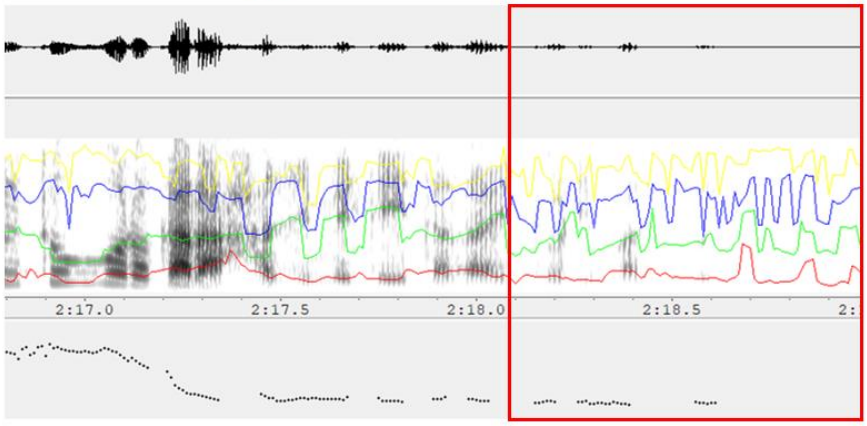


図13　　WaveSurferの表示例（語尾上がり評点１）

・評点１の例

・発話末尾の数フレームのピッチの変化量（Δ）が上昇

※但し、上げ下げの正誤は文意によるので、文意を考慮して分析要

発話内容：ハイレートなどはあんまりご検討はされてないでしょうか↗

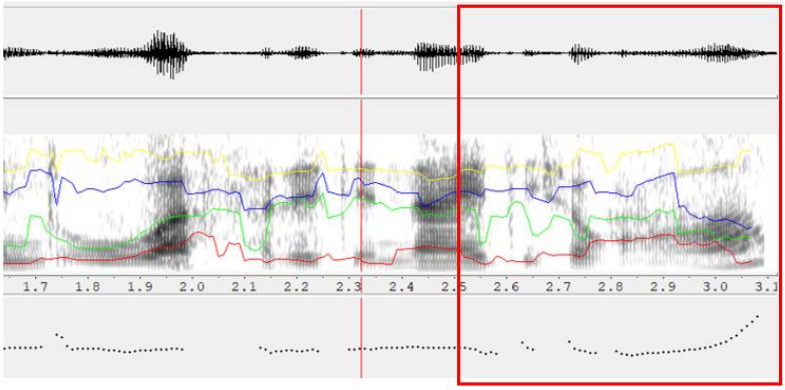


図14　　WaveSurferの表示例（語尾下がり評点１）

・評点1の例

・発話末尾の数フレームのピッチの変化量（Δ）が下降

※但し、上げ下げの正誤は文意によるので、文意を考慮して分析要

発話内容：認定されているようではございますが↘

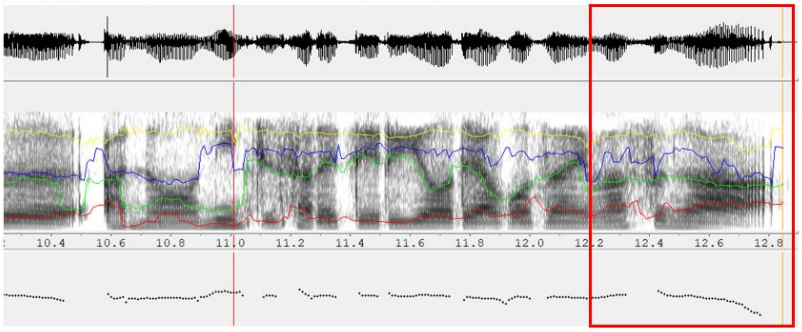
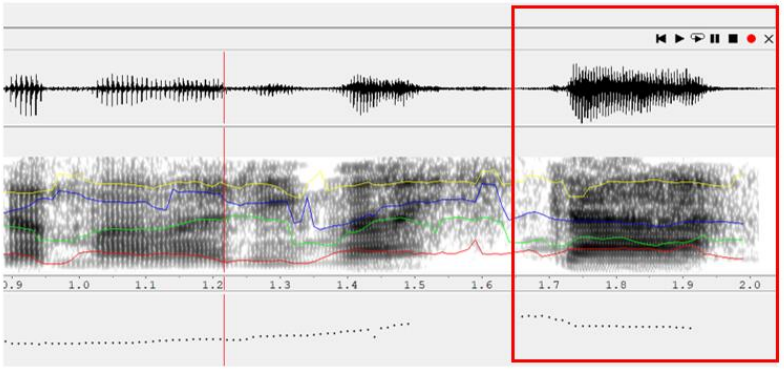


図15　　WaveSurferの表示例（語尾強さ評点１）

・評点１の例
・発話末尾数フレームの音量が全体と比較して大きい

発話内容：【担当者名】が承りました



WaveSurferを活用し、語尾の各評価項目については発話末尾から0.3秒～1秒間のピッチや音量の大小が影響していることが確認できた。特に、語尾跳ねと語尾上がりの違いは、語尾跳ねの方がより短時間での変化が激しいこともわかった。よって、自動推定においては、発話末尾から0.3秒～1秒間の音声データを切り取り、その中での特徴量を作成しモデルを構築することとした。

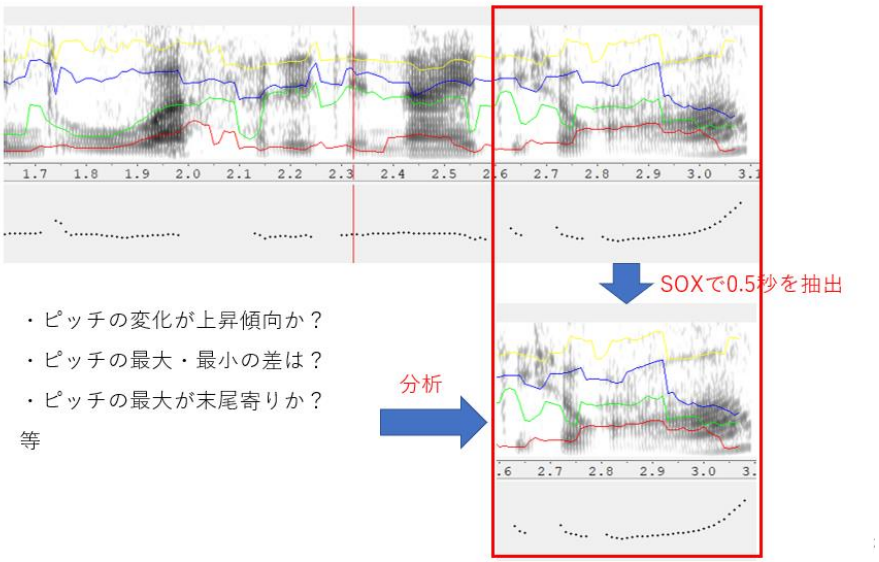
発話末尾の音声データの作り方は、SoXというツールを使用した。SoX(Sound eXchange)とは、サウンドプロセッシングプログラムである。さまざまな形式のオーディオファイルを他の形式に変換できる。SoX(Sound eXchange)はクロスプラットフォームのコマンドラインユーティリティで、さまざまな形式のオーディオファイルを他の形式に変換でき、オーディオ形式変換機能に加えて、「サウンドエフェクト適用機能」「オーディオファイル再生機能」「オーディオファイル録音機能」なども備えている。SoXの主な機能は以下である。

・フォーマット変換機能
・サンプリング変換機能
・オーディオエフェクト機能
・ファイル結合、分割機能
等

上記のSoXを使用し、その機能の内でファイル分割機能により末尾から時間指定でファ

イルを切り出し、その切り出したファイルから特徴量を作成する。例えば、図16のように、評価項目８：語尾上がりであれば、発話末尾0.5秒の音声データをSoXで切り出し、ピッチの変化が上昇傾向か？、ピッチの最大・最小の差は？、ピッチの最大が末尾寄りか？、等の特徴量を作成する。

図16　　SOXによる末尾0.5秒の抽出・分析



切り出した末尾の音声ファイルから、各評価項目に対して分析方針に従った特徴量を作成した。基本的には末尾音声ファイルの音量やF0（ピッチ）の変化を捉えられるような特徴量となっている。各評価項目に対する特徴量は表11の通りである。この特徴量を使用して、評点１，２，３の推定モデルを構築した。

表11　語尾系評価項目に対する特徴量

No	評価項目	特徴量
5	語尾跳ね	発話末尾0.3秒：F0平均、F0最小、F0最大、F0最大−F0最小、F0Δ平均、F0minPos、F0maxPos、F0maxPos−F0minPos
6	語尾消え	発話末尾0.5秒：音量平均、音量最小、音量最大、音量平均/発話全体音量平均
7	語尾伸び	発話末尾1秒　：F0平均、F0最小、F0最大、F0最大−F0最小、F0Δ平均、F0Δ分散
8	語尾上がり	発話末尾0.5秒：F0平均、F0最小、F0最大、F0最大−F0最小、F0Δ平均、F0minPos、F0maxPos、F0maxPos−F0minPos
9	語尾下がり	発話末尾0.5秒：F0平均、F0最小、F0最大、F0最大−F0最小、F0Δ平均、F0minPos、F0maxPos、F0maxPos−F0minPos
10	語尾強さ	発話末尾0.5秒：音量平均、音量最小、音量最大、音量平均/発話全体音量平均

F0maxPos、F0minPos：F0の最大、最小が発生している箇所（フレーム単位）

各評価項目に対して構築したモデルのテスト結果は以下である。機械学習手法としてはAdaboostを使用した。Adaboostはアンサンブル学習の一種で、決定木やロジスティック回帰、サポートベクターマシン（SVM）といった教師ありの機械学習手法と組み合わせて使えるメタアルゴリズムである。機械学習手法と組み合わせることでパフォーマン

スを改善することができる。アンサンブル学習とは複数の機械学習モデル（弱学習器）を組み合わせて一つの学習モデル（強学習器）を生成する手法のことである。今回は、Adaboostと決定木の組み合わせで実施した。検証方法はk-分割交差検証法（k-Fold-CV）で実施した。その実施結果は表12である。語尾伸びの正解率と、各評価項目における評点１の再現率に課題があるものの、ある程度自動推定できる結果を残すことができた。語尾系の評価項目に対する今後の課題としては、今回は発話末尾0.3秒、0.5秒と決め打ちで音声データを切り出し、特徴量を作成したが、例えば語尾上がりのケースで言えば、語尾上がりが発生している正確な時間を音声データから把握することができれば、その時間における変化を捉えることにより精度向上を図ることが可能かもしれない。

表12　　語尾系評価項目の実験結果

語尾跳ね

正解＼推定結果	1	2	3
1	8	15	11
2	13	71	119
3	18	110	1871

再現率	適合率	F値
0.235	0.205	0.219
0.35	0.362	0.356
0.936	0.935	0.936

正解率
87.2

語尾消え

正解＼推定結果	1	2	3
1	17	22	7
2	21	52	66
3	11	67	1973

再現率	適合率	F値
0.37	0.347	0.358
0.374	0.369	0.371
0.962	0.964	0.963

正解率
91.3

語尾伸び

正解＼推定結果	1	2	3
1	4	20	52
2	40	322	387
3	30	382	999

再現率	適合率	F値
0.053	0.054	0.053
0.43	0.445	0.43
0.708	0.695	0.708

正解率
59.2

語尾上がり

		推定結果		
		1	2	3
正解	1	7	13	4
	2	11	112	69
	3	1	77	1942

再現率	適合率	F値
0.292	0.368	0.326
0.583	0.554	0.569
0.961	0.964	0.963

正解率
92.1

語尾下がり

		推定結果		
		1	2	3
正解	1	5	17	1
	2	7	270	126
	3	0	94	1716

再現率	適合率	F値
0.217	0.417	0.286
0.67	0.709	0.689
0.948	0.931	0.94

正解率
89

語尾強さ

		推定結果		
		1	2	3
正解	1	34	36	5
	2	40	319	167
	3	9	151	1475

再現率	適合率	F値
0.453	0.41	0.43
0.606	0.63	0.618
0.902	0.896	0.899

正解率
81.7

3.3.2.1 文意による層別分析

ここまで、語尾系の評価項目について実験してきた。全体的に評点１の再現率が低く、精度改善が求められる。そこで、モデルの精度向上策として、文意による層別分析を行い、文意を考慮したモデルにより精度改善できないか実験してみた。文意とは、音声データに対して音声認識を通して出力されたテキストデータに対してラベリングされた以カテゴリ情報のことである。図17にて文意カテゴリ内容を示す。

図17　　文意のカテゴリ内容

【文意A】
①挨拶
②受け止め
③確認
④案内･説明

【文意B】
①明るさ･前向き　　(明るくリードをもって応対したい場面。歓迎姿勢)
②共感･心配　　(相手に合わせて感情を伝えたい場面。一緒に立ち止まる姿勢)
③責任･慎重　　(問題解決のために対等な関係を保ち、真摯な姿勢を見せたい場面。解決姿勢)
④謝罪･丁重　　(低姿勢に相手を十分に気遣い応対をしたい場面。お詫び姿勢)

そこで、文意と語尾の評価項目に関係があるかを、評価項目８「語尾上がり」を例に、末尾0.5秒の音量分散・F0分散の関係を図示してみた。文意Aと音量分散・F0分散の関係

が図18、文意Bと音量分散・F0分散の関係が図19、文意Aと音量分散Δ・F0分散Δの関係が図20、文意Bと音量分散Δ・F0分散Δの関係が図21である。

図18　　文意Aにおける評点と音量、F0分散の散布図

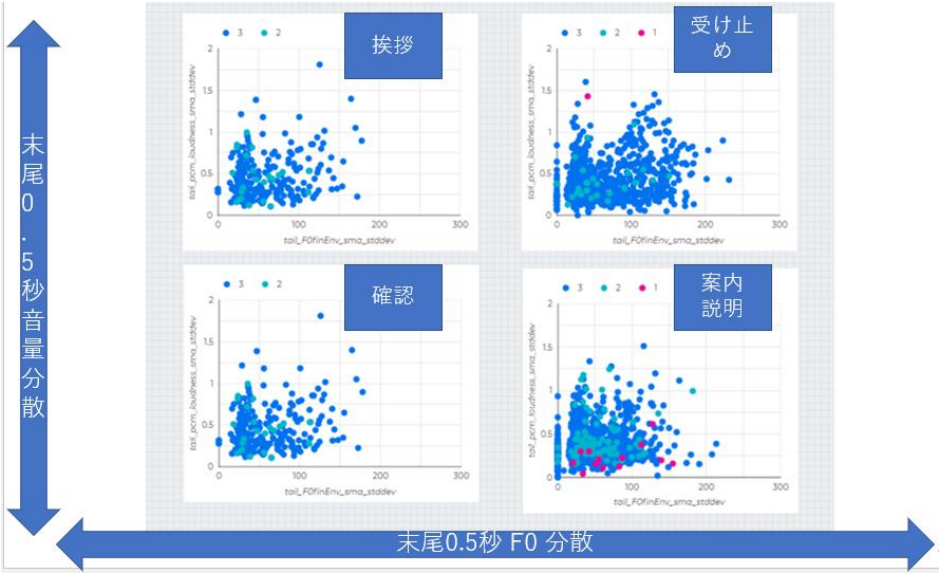


図19　　文意Bにおける評点と音量、F0分散の散布図

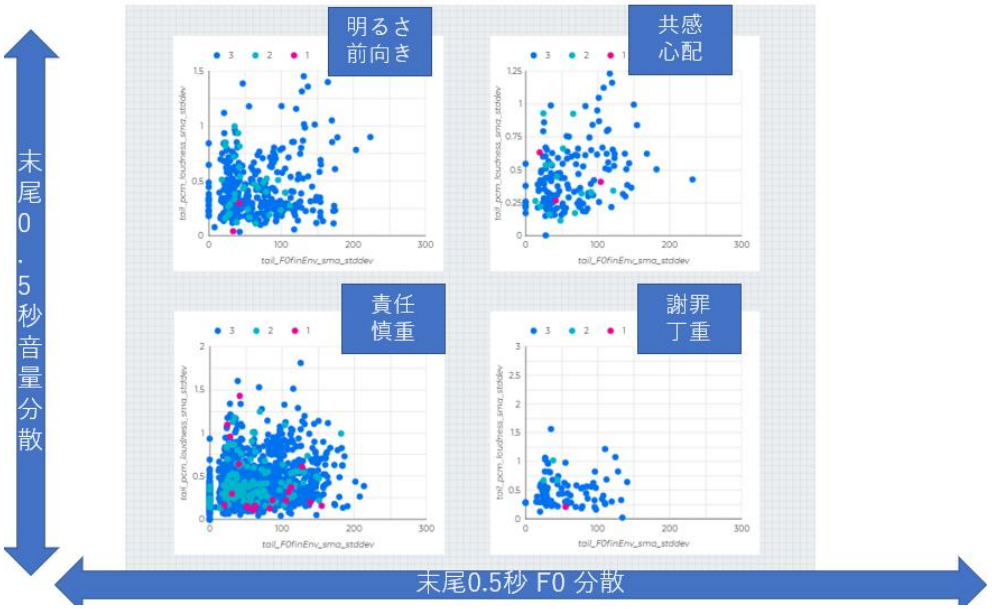


図20　　文意Aにおける評点と音量Δ、F0分散Δの散布図

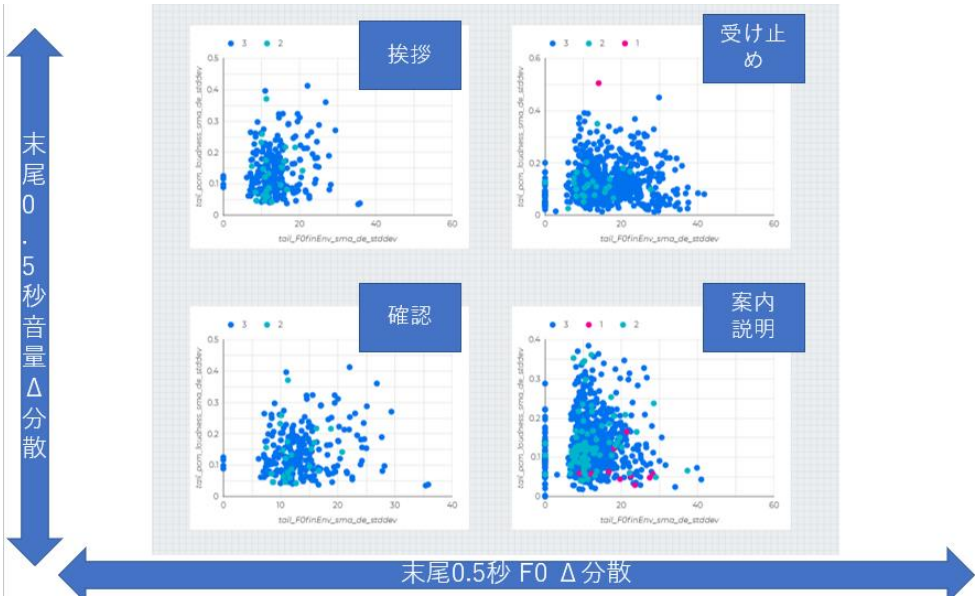
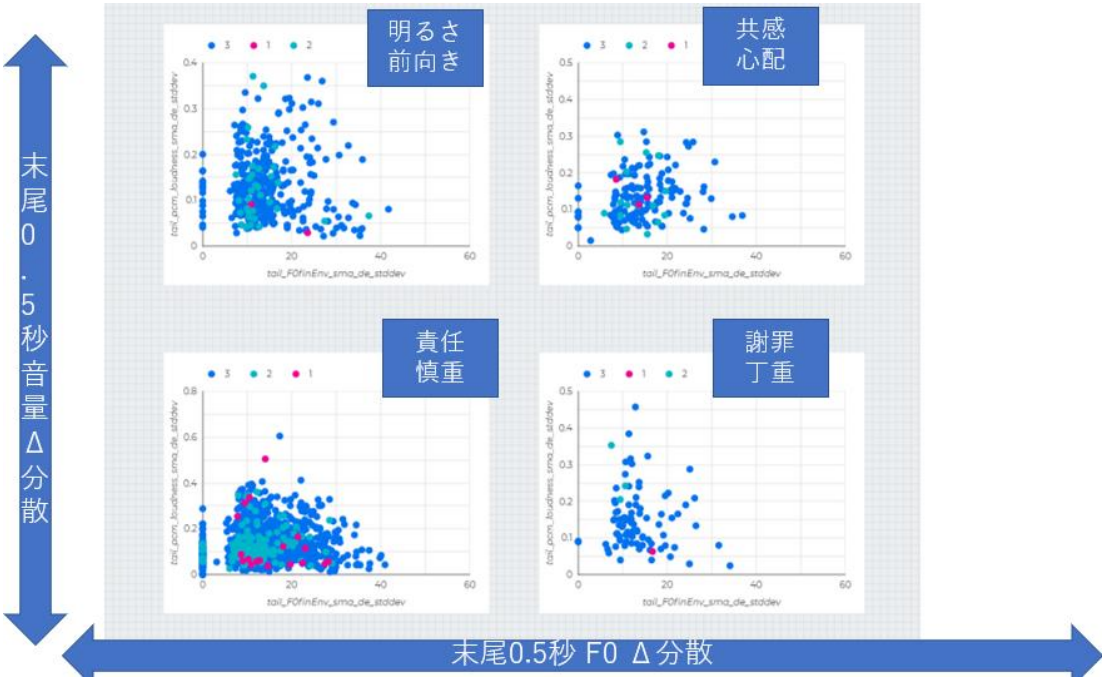


図21　　文意Bにおける評点と音量Δ、F0分散Δの散布図



このように文意と末尾0.5秒の各特徴量にはある程度の関係があることが見て取れる。

そこで、精度向上として、入力特徴量に文意A,Bのラベリング情報を加えて実験してみることにした。機械学習手法としてはAdaboostを使用し、弱学習器として決定木との組み合わせで実施した。検証方法はk-分割交差検証法（k-Fold-CV）で実施した。その実験結果は表13から表18である。

表13　　評価項目８語尾上がり、文意ABラベル有無比較

文意なし		推定結果		
		1	2	3
正解	1	7	13	4
	2	11	112	69
	3	1	77	1942

再現率	適合率	F値
0.292	0.368	0.326
0.583	0.554	0.569
0.961	0.964	0.963

正解率
92.1

文意AB		推定結果		
		1	2	3
正解	1	4	16	4
	2	3	99	90
	3	0	45	1975

再現率	適合率	F値
0.167	0.571	0.258
0.516	0.619	0.563
0.978	0.955	0.966

正解率
92.9

表14　　評価項目８語尾上がり、文意Aラベル有無比較

文意なし		推定結果		
		1	2	3
正解	1	7	13	4
	2	11	112	69
	3	1	77	1942

再現率	適合率	F値
0.292	0.368	0.326
0.583	0.554	0.569
0.961	0.964	0.963

正解率
92.1

文意A		推定結果		
		1	2	3
正解	1	7	16	1
	2	7	112	73
	3	0	51	1969

再現率	適合率	F値
0.292	0.5	0.368
0.583	0.626	0.604
0.975	0.964	0.969

正解率
89

表15　　評価項目８語尾上がり、文意Bラベル有無比較

文意なし

		推定結果		
		1	2	3
正解	1	7	13	4
	2	11	112	69
	3	1	77	1942

再現率	適合率	F値
0.292	0.368	0.326
0.583	0.554	0.569
0.961	0.964	0.963

正解率
92.1

文意B

		推定結果		
		1	2	3
正解	1	8	10	6
	2	8	102	82
	3	0	51	1969

再現率	適合率	F値
0.333	0.5	0.4
0.531	0.626	0.575
0.975	0.957	0.966

正解率
92.9

表16　評価項目８語尾下がり、文意ABラベル有無比較

文意なし

		推定結果		
		1	2	3
正解	1	5	17	1
	2	7	270	126
	3	0	94	1716

再現率	適合率	F値
0.217	0.417	0.286
0.67	0.709	0.689
0.948	0.931	0.94

正解率
89

文意AB

		推定結果		
		1	2	3
正解	1	5	18	0
	2	5	274	124
	3	0	94	1716

再現率	適合率	F値
0.217	0.5	0.303
0.68	0.71	0.695
0.948	0.933	0.94

正解率
89.2

表17　　評価項目9語尾下がり、文意Aラベル有無比較

文意なし

		推定結果		
		1	2	3
正解	1	5	17	1
	2	7	270	126
	3	0	94	1716

再現率	適合率	F値
0.217	0.417	0.286
0.67	0.709	0.689
0.948	0.931	0.94

正解率
89

文意A

		推定結果		
		1	2	3
正解	1	5	17	1
	2	5	282	116
	3	0	87	1723

再現率	適合率	F値
0.217	0.5	0.303
0.7	0.731	0.715
0.952	0.936	0.944

正解率
89.8

表18　　評価項目9語尾下がり、文意Bラベル有無比較

文意なし

		推定結果		
		1	2	3
正解	1	5	17	1
	2	7	270	126
	3	0	94	1716

再現率	適合率	F値
0.217	0.417	0.286
0.67	0.709	0.689
0.948	0.931	0.94

正解率
89

文意B

		推定結果		
		1	2	3
正解	1	5	17	1
	2	5	280	119
	3	0	97	1713

再現率	適合率	F値
0.217	0.556	0.313
0.695	0.711	0.703
0.946	0.935	0.94

正解率
89.3

3.3.2.2 単語クラスによる層別分析

前節では、語尾系評価項目に対して文意による層別分析、及び文意ラベルを特徴量として加えてモデルの精度向上を試みたが大きな改善には至らなかった。ここでは、さらに語尾の特定単語クラスに着目し精度向上を試みる。着目する語尾の特定単語クラスを表19に示す。この単語クラスは、ビーウィズ社から提供されているコールセンター音声データで実際に発生したオペレーター発話の語尾の単語を整理したものである。

表19　　語尾の特定単語クラス

大項目	中項目
です t7	です
	ですね
	ですよね
	ですよ
	ですか
	ですかね
ます t8	ます
	ますね
	ますよね
	ました
	ません
	おります
	ございます
	いたします
	ませ
	ますか
	いたしますか
しょうか t9	しょうか
	ましょうか
	でしょうか
	ますでしょうか
	ございましょうか

語尾止め t10	ですが
	ですと
	ですけど
	でして
	まして
	ますと
	ので
	けれども
	であれば
	でしたら
	ですから
	ですとか
他 t11	下さい/ください

まず、各評価項目の評点１が多く発生している特定単語クラスに限定してモデルを構築してみる。例えば、表20のように、評価項目８「語尾上がり」での場合では、語尾止めのケースで評点１の件数が多いため、語尾止めのデータに限定しモデルを構築する。その他の評価項目も同様に評点１の発生が多い単語クラスのデータに限定してモデルを構築する。語尾のラベルについては、ビーウィズ社からのラベル情報には存在しないため、実験のため自身でラベルを付与した。ラベルの付与方法は、ビーウィズ社から提供されている音声データのテキスト情報から末尾10語の内容を抽出し、特定単語で検索しマッチした対象データにラベルを付与した。例として、表21に評価項目8の語尾上がりで評点1が多い単語クラスが発生している発話内容を示す。

表20　　評価項目８語尾上がり、評点１が多い単語（赤枠）

大項目	中項目
です t7	です
	ですね
	ですよね
	ですよ
	ですか
	ですかね
ます t8	ます
	ますね
	ますよね
	ました
	ません
	おります
	ございます
	いたします
	ませ
	ますか
	いたしますか
しょうか t9	しょうか
	ましょうか
	でしょうか
	ますでしょうか
	ございましょうか

語尾止め t10	ですが
	ですと
	ですけど
	でして
	まして
	ますと
	ので
	けれども
	であれば
	でしたら
	ですから
	ですとか
他 t11	下さい/ください

表21　　評価項目８語尾上がり、評点１の具体例

文意A	文意B	Id	Recogn	RecognitionText2	合計	【他】	下さい	ですとか	【語尾上	ですと
④案内・説	③責任・恒	2	そうしま	と思うんですけれども	54				1	
④案内・説	③責任・恒	2	あのご使	ことはあると思うので	41				1	
④案内・説	③責任・恒	2	あと残り	だったんですけれども	43				1	
④案内・説	①明るさ・	2	色々種類	せて頂きたいんですが	51				1	
④案内・説	③責任・恒	2	あとはへ	例えば吸引力ですとか	45			1	1	
④案内・説	③責任・恒	2	そうする	掃除機でございますと	48				1	

特定単語クラスに限定したケースでの評価項目５から10のそれぞれのモデル構築結果は表22から表27である。機械学習手法としてはAdaboostを使用し、弱学習器として決定木との組み合わせで実施した。検証方法はk-分割交差検証法（k-Fold-CV）で実施した。多くの評価項目で評点１の再現率を改善することができた。

表22　　評価項目８語尾上がり、単語限定

トータル

正解＼推定結果	1	2	3
1	7	13	4
2	11	112	69
3	1	77	1942

再現率	適合率	F値
0.292	0.368	0.326
0.583	0.554	0.569
0.961	0.964	0.963

正解率
92.1

語尾止め

正解＼推定結果	1	2	3
1	3	4	2
2	2	22	14
3	2	8	235

再現率	適合率	F値
0.33	0.429	0.375
0.579	0.647	0.611
0.959	0.883	0.948

正解率
89

表23　　評価項目９語尾下がり、単語限定

トータル

正解＼推定結果	1	2	3
1	5	17	1
2	7	270	126
3	0	94	1716

再現率	適合率	F値
0.217	0.417	0.286
0.67	0.709	0.689
0.948	0.931	0.94

正解率
89

ます

正解＼推定結果	1	2	3
1	7	14	0
2	9	155	59
3	2	53	882

再現率	適合率	F値
0.33	0.389	0.359
0.695	0.698	0.697
0.941	0.937	0.939

正解率
88.3

表24　　評価項目10語尾強さ、単語限定

トータル

		推定結果		
		1	2	3
正解	1	34	36	5
	2	40	319	167
	3	9	151	1475

再現率	適合率	F値
0.453	0.41	0.43
0.606	0.63	0.618
0.902	0.896	0.899

正解率
81.7

ですね

		推定結果		
		1	2	3
正解	1	7	5	0
	2	12	23	8
	3	0	10	21

再現率	適合率	F値
0.583	0.368	0.452
0.535	0.605	0.568
0.677	0.724	0.7

正解率
59.3

表25　　評価項目5語尾跳ね、単語限定

トータル

		推定結果		
		1	2	3
正解	1	8	15	11
	2	13	71	119
	3	18	110	1871

再現率	適合率	F値
0.235	0.205	0.219
0.35	0.362	0.356
0.936	0.935	0.936

正解率
87.2

です系

		推定結果		
		1	2	3
正解	1	8	3	5
	2	5	10	12
	3	3	12	180

再現率	適合率	F値
0.5	0.5	0.452
0.37	0.4	0.568
0.923	0.914	0.7

正解率
83.1

表26　　評価項目6語尾消え、単語限定

トータル

		推定結果		
		1	2	3
正解	1	17	22	7
	2	21	52	66
	3	11	67	1973

再現率	適合率	F値
0.37	0.347	0.358
0.374	0.369	0.371
0.962	0.964	0.963

正解率
91.3

ます系

		推定結果		
		1	2	3
正解	1	13	13	7
	2	6	43	44
	3	8	35	1008

再現率	適合率	F値
0.394	0.481	0.452
0.462	0.473	0.568
0.959	0.952	0.7

正解率
90.3

表27　　評価項目7語尾伸び、単語限定

語尾伸び

		推定結果		
		1	2	3
正解	1	4	20	52
	2	40	322	387
	3	30	382	999

再現率	適合率	F値
0.053	0.054	0.053
0.43	0.445	0.43
0.708	0.695	0.708

正解率
59.2

ます系

		推定結果		
		1	2	3
正解	1	14	23	30
	2	31	224	203
	3	14	214	443

再現率	適合率	F値
0.209	0.237	0.222
0.489	0.486	0.487
0.655	0.655	0.658

正解率
56.9

ここまでは、各評価項目に対する評点１の件数が多い単語クラスに限定してモデルを構築した。ここからは、特定単語クラスに限定せずに、特定単語クラスのラベル情報を特徴量として付与してモデルを構築する。ラベルの付与単位は表19の特定単語表にある大項目単位のラベル情報である。（です、ます、しょうか、語尾止め、他、5次元追加）

ラベルの付与有無による各評価項目に対する実験結果は表28から表33である。機械学習手法としてはAdaboostを使用し、弱学習器として決定木との組み合わせで実施した。検証方法はk-分割交差検証法（k-Fold-CV）で実施した。多くの評価項目で評点１の再現率を向上させることができた。

表28　　評価項目8語尾上がり、単語ラベル

ラベルなし

		推定結果		
		1	2	3
正解	1	7	13	4
	2	11	112	69
	3	1	77	1942

再現率	適合率	F値
0.292	0.368	0.326
0.583	0.554	0.569
0.961	0.964	0.963

正解率
92.1

ラベルあり

		推定結果		
		1	2	3
正解	1	9	13	2
	2	10	110	72
	3	1	66	1953

再現率	適合率	F値
0.375	0.45	0.409
0.573	0.582	0.577
0.967	0.963	0.965

正解率
92.6

表29　　評価項目9語尾上がり、単語ラベル

ラベルなし

		推定結果		
		1	2	3
正解	1	5	17	1
	2	7	270	126
	3	0	94	1716

再現率	適合率	F値
0.217	0.417	0.286
0.67	0.709	0.689
0.948	0.931	0.94

正解率
89

ラベルあり

		推定結果		
		1	2	3
正解	1	9	13	1
	2	15	267	121
	3	0	125	1685

再現率	適合率	F値
0.391	0.375	0.383
0.669	0.663	0.661
0.932	0.931	0.932

正解率
87.7

表30　　評価項目10語尾強さ、単語ラベル

ラベルなし

		推定結果		
		1	2	3
正解	1	34	36	5
	2	40	319	167
	3	9	151	1475

再現率	適合率	F値
0.453	0.41	0.43
0.606	0.63	0.618
0.902	0.896	0.899

正解率
81.7

ラベルあり

		推定結果		
		1	2	3
正解	1	32	38	5
	2	34	335	157
	3	2	150	1483

再現率	適合率	F値
0.427	0.471	0.448
0.637	0.641	0.639
0.907	0.902	0.904

正解率
82.7

表31　　評価項目5語尾跳ね、単語ラベル

ラベルなし

		推定結果		
		1	2	3
正解	1	8	15	11
	2	13	71	119
	3	18	110	1871

再現率	適合率	F値
0.235	0.205	0.219
0.35	0.362	0.356
0.936	0.935	0.936

正解率
87.2

ラベルあり

		推定結果		
		1	2	3
正解	1	10	13	11
	2	12	77	114
	3	9	80	1910

再現率	適合率	F値
0.294	0.332	0.308
0.379	0.463	0.413
0.955	0.939	0.947

正解率
89.3

表32　　評価項目6語尾消え、単語ラベル

ラベルなし

		推定結果		
		1	2	3
正解	1	17	22	7
	2	21	52	66
	3	11	67	1973

再現率	適合率	F値
0.37	0.347	0.358
0.374	0.369	0.371
0.962	0.964	0.963

正解率
91.3

ラベルあり

		推定結果		
		1	2	3
正解	1	18	20	8
	2	20	55	64
	3	7	73	1971

再現率	適合率	F値
0.391	0.4	0.396
0.396	0.372	0.382
0.961	0.965	0.963

正解率
91.4

表33　　評価項目7語尾伸び、単語ラベル

ラベルなし

		推定結果		
		1	2	3
正解	1	4	20	52
	2	40	322	387
	3	30	382	999

再現率	適合率	F値
0.053	0.054	0.053
0.43	0.445	0.43
0.708	0.695	0.708

正解率
59.2

ラベルあり

		推定結果		
		1	2	3
正解	1	5	32	39
	2	42	357	350
	3	34	383	994

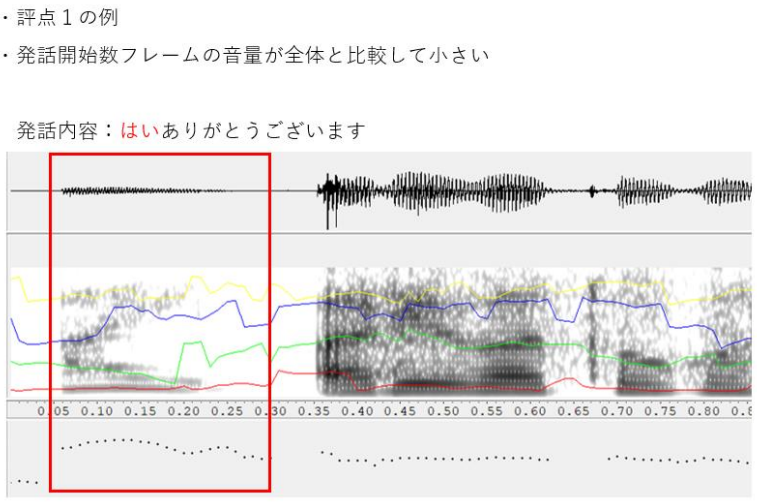
再現率	適合率	F値
0.066	0.062	0.064
0.477	0.462	0.469
0.704	0.719	0.712

正解率
60.6

3.3.3 語頭の自動推定

ここでは、語頭の自動推定について述べる。3.3.2節の語尾の自動推定と同様に、WaveSurferを使って、評価項目に対する評点１のwaveファイルを入力に音声波形、スペクトログラム、F0の時間変化を図22の通り表示してみた。語頭の音量がその後の音量に比べ小さいことがわかる。

図22　　語頭の評点１の例



よって、語頭の自動推定に使う特徴量として以下を考えた。

・発話開始0.5秒音量平均
・発話開始0.5秒音量最小
・発話開始0.5秒音量最大、
・発話開始0.5秒音量平均/発話全体音量平均

特徴量の生成の仕方は3.3.2節の語尾自動推定に用いた特徴量と同様に、SOXで発話開始0.5秒のWAVEファイルを切り出し、そのファイルを入力にOpensmileでIS10特徴量を抽出し、上記特徴量を作成した。この特徴量を使用した自動推定結果は表34である。機械学習手法としてはAdaboostを使用し、弱学習器として決定木との組み合わせで実施した。検証方法はk-分割交差検証法（k-Fold-CV）で実施した。評点１の再現率、適合率が低い結果となった。これは語尾系評価項目と同様に、発話開始0.5秒の情報で特徴量を作っているため、正確な語頭の時間に対する特徴量を捉えられていないのが原因であると考える。語尾と同様に、アライメントによる音素を分析し、正確な語頭発生時間を判定し特徴量を生成することができれば精度が上がる可能性はあるが、今後の課題である。

表34　　語頭の自動推定結果

		推定結果		
		1	2	3
正解	1	6	17	9
	2	17	103	153
	3	21	135	850

再現率	適合率	F値
0.188	0.136	0.158
0.377	0.404	0.39
0.845	0.94	0.842

正解率
88.1

3.3.4　滑舌の自動推定

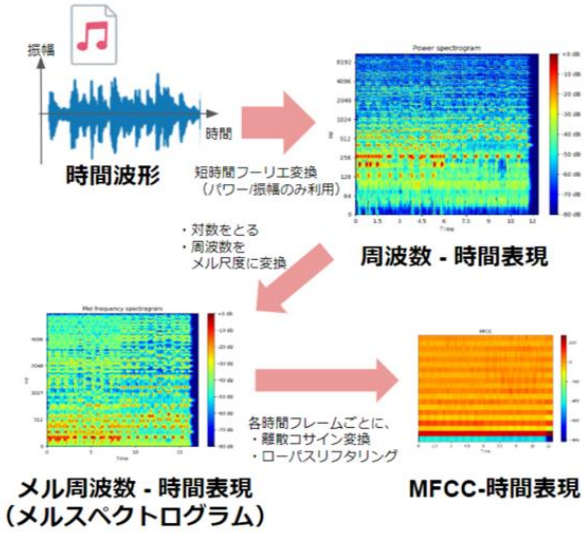
ここでは滑舌の自動推定について述べる。滑舌（かつぜつ）とは、舌の動きを滑らかにしてはっきりと聞き取りやすい発音をしている状態のことである。一般的には、言葉の発音がはっきりとしていて聞き取りやすい人のことを『滑舌が良い』、逆に発音があいまいで聞き取りづらい人のことを『滑舌が悪い』と呼ばれる（舌足らずと呼ぶ人もいる）。滑舌が悪いと、歌や会話でも何を言っているのか分からないと思われてしまうため、自分の思いや意見がストーレートに伝わらないというデメリットがある。実際、何度も聞き返されてしまうという人も少なくない。

このように、滑舌の良し悪しは、言葉の発音がはっきりとしていて聞き取りやすいか否かであるか、ということであるため、これは音響特徴量であるMFCC（メルフィルターケプストラム）で捉えられるのではないかと考えた。MFCCは音声認識や音楽ジャンル分類などで使われる特徴量であり、人間の聴覚特性を考慮した周波数スペクトルの概形（包絡線）を表している。MFCCの変化の大小により、滑舌の良し悪しを判別できるのではないかと考えた。MFCCの算出の流れの概要は以下である。また、その流れを図23にて示す。

1. 時間信号をフーリエ変換して周波数信号を算出
2. 周波数信号のパワー or 振幅成分を抽出し、対数をとる
    - 人間の音量（音圧）に対する知覚は対数的であるため
3. ↑ をメル尺度（メル周波数） に変換し、メル周波数スペクトルを抽出
    - 人間の音高（周波数）に対する聴感特性を考慮
    - 各メル周波数に対応した三角窓を使って抽出
4. メル周波数スペクトルをケプストラム分析
    - メル周波数スペクトルを離散コサイン変換し、メル周波数ケプストラム (MFC) を抽出
    - MFC の低次元成分を取り出す（ローパスリフタリング）

- ＝メル周波数スペクトルの概形を取り出す。つまりメル周波数スペクトルの高周波数成分を無視
- 0 次元目も無視（直流成分であるため）

図23　　MFCC算出の流れ



上記の手順でMFCCの時間表現を算出し、「評価項目：滑舌が悪くないか」について、評点１（滑舌が悪い）と評点３（滑舌が良い）のデータのMFCC時間変化ヒートマップを表示し比較してみた。その比較結果が図24である。評点１（滑舌が悪い）のMFCCは時間変化が小さく、評点３（滑舌が良い）はMFCCの時間変化が大きいのがヒートマップの比較で明らかである。そこで、MFCCの時間変化を捉えられる特徴量として、表35の通りMFCCの1次から24次の各平均、分散、1階微分Δ分散を考えた。

図24　　滑舌良し悪しに対するMFCCのヒートマップ

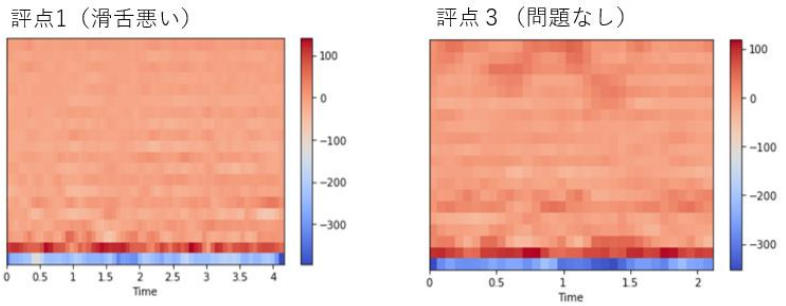


表35　　滑舌悪さに対する特徴量

評価項目	特徴量
滑舌悪さ	mfcc1次～24次の各平均、mfcc1次～24次の各分散、mfcc1次～24次の各Δ分散

実行結果は表36である。機械学習手法としてはAdaboostを使用し、弱学習器として決定木との組み合わせで実施した。検証方法はk-分割交差検証法（k-Fold-CV）で実施した。件数が少ない評点１（滑舌が悪い）も含め、全体的に高い精度で推定できており、自動化に向けて大きな可能性があることがわかった。

表36　　滑舌が悪くないか実験結果

		推定結果		
		1	2	3
正解	1	9	6	1
	2	3	602	71
	3	0	107	516

再現率	適合率	F値
0.563	0.75	0.643
0.891	0.842	0.866
0.828	0.878	0.852

正解率
85.7

3.3.5 抑揚の自動推定

ここでは抑揚について述べる。抑揚とは、話すときの音声や文章などで、調子を上げたり下げたりすることで、英語ではイントネーションと呼ばれる。音楽では無伴奏で歌われる単声聖歌の始めの部分である。また、正確な音の高さとも言う。

言語学では、話しことば、朗読などにおける声の高さの時間的変化をいう。句や文の終結と連続とを区別し、平叙文、断定文、命令文、疑問文の別を示し、また、喜怒哀楽など話者の感情を表出する。とくに統語構造、話の焦点などを明示するため、内容理解への影響が大きい。一般に、文の末尾は低く、次の句へ連続する場合の句末はやや高い。疑問文では文末が上昇する。上昇の度合いには、言語差および方言差があり、英語より日本語のほうが上昇幅が少なく、東京方言より近畿方言のほうがより少ない傾向がある。疑問文における末尾の上昇は各言語に共通する特徴と考えられているが、言語により異なる場合も少なくない。英語では、疑問詞が先行する文の末尾は上昇せず、Yes, Noで答えられる疑問文は末尾が上昇する。この種の言語は多いが、ロシア語では、「マーマ・ドーマ？」Мама　дома?（お母さんはお家？）の場合、доで急に上昇し、末尾で急に下降する。

いわゆる強弱アクセントの言語では、アクセントは強さ、イントネーションは高さの変化であるとされる。最近の実験結果によれば、アクセントのもっとも重要な成分は高さであり、強さ、長さおよび音質の変化がこれに伴う。イントネーションも高さの変化が主体であるが、これとともに、強さ、長さおよび音質の変化が伴う場合がある。たとえば、平叙文と疑問文とは末尾の基本周波数の変化（高さの変化）であり、喜び、怒り、悲しみなどの感情表現は、高さの変化だけではなく、音質と長さおよび強さの変化を伴う。感情表現における音響的特徴においても、人間としての普遍性があるとともに、言語による差異があり、英語より日本語のほうが変化が少ない。イントネーションに関し

ても、アクセントと同様に、生成（発話）と知覚（聞こえ）の両面から検討すべき点が多い。

このように、抑揚は声の高さの変化で表されるため、評価項目「抑揚が極端ではないか」「抑揚が小さすぎないか」についての判定モデルとしてはF0分散値、F0最大値、F0最小値、F0Δ分散値を特徴量として考えた。この特徴量はIS10（opensmile特徴量）に含まれているため、当初はIS10特徴量全入力でSVMにより実施した。その時の精度としては評点１の再現率が全くなかったため、その後、上記の４つの特徴量へ限定し、機械学習手法としてはAdaboostを使用し、弱学習器として決定木との組み合わせで実施し直した。検証方法はk-分割交差検証法（k-Fold-CV）で実施した。全体の正解率及び評点１の再現率の精度を向上させることができた。その実行の比較結果は表37、表38である。結果として評点１の精度に課題が残る結果となった。これは、イントネーション句（抑揚）に対するF0の分散値だけでは、アクセントパターンのミスマッチがあった場合に正しく評価できてない可能性がある。

例：発話全体の抑揚は普通だが、短いアクセント句内の抑揚が極端なケース

発話内容に対する標準的なアクセント句の区切りやアクセントパターンと比べ、オペレーター音声がどの程度ミスマッチしているか判定ができれば精度の向上が見込まれるが、これは今後の課題である。

表37　　抑揚が極端ではないか実験結果

SVM+opensmile

		推定結果		
		1	2	3
正解	1	0	10	0
	2	3	37	47
	3	1	36	1181

再現率	適合率	F 値
0	0	0
0.452	0.446	0.435
0.97	0.962	0.966

正解率
92.6

SVM+3つの特徴量

		推定結果		
		1	2	3
正解	1	0	6	5
	2	0	10	77
	3	0	4	1214

再現率	適合率	F 値
0	0	0
0.115	0.526	0.189
0.997	0.937	0.966

正解率
93

決定木+３つの特徴量

		推定結果		
		1	2	3
正解	1	3	7	1
	2	4	37	46
	3	2	18	1198

再現率	適合率	F 値
0.3	0.333	0.316
0.425	0.597	0.497
0.984	0.963	0.937

正解率
94.1

表38　　抑揚が小さすぎないか実験結果

SVM+opensmile

		推定結果		
		1	2	3
正解	1	0	1	1
	2	3	56	97
	3	0	87	1073

再現率	適合率	F 値
0	0	0
0.366	0.389	0.377
0.925	0.916	0.921

正解率
85.8

SVM+3つの特徴量

		推定結果		
		1	2	3
正解	1	0	0	2
	2	0	39	111
	3	0	111	1049

再現率	適合率	F 値
0	0	0
0.255	0.26	0.257
0.904	0.827	0.904

正解率
82.7

決定木+３つの特徴量

		推定結果		
		1	2	3
正解	1	0	2	0
	2	1	48	104
	3	0	64	1096

再現率	適合率	F 値
0	0	0
0.314	0.421	0.36
0.945	0.913	0.929

正解率
86.9

3.3.6　話速の自動推定

ここでは話速の自動推定について述べる。一般に、話速は発話のモーラ数をその発話の持続時間で割った値が用いられる。また、アナウンサーの話速としては、1 分間のモーラ数によって定義されることが多い。例えば、NHK のニュースアナウンスの話速は、約 330 モーラ/分 であることが知られている。モーラ、モラ（mora）とは、音韻論上、一定の時間的長さをもった音の分節単位である。古典詩における韻律用語であるラテン語の mŏra ['mɔra]（モラ）の転用（日本語における「モーラ」という表記はラテン語からの借用語の英語の mora ['mɔːɹə] からの音訳であり、「モラ」という表記はラテン語からの音訳）。拍（はく）とも訳される。音韻の構造によって定められる音節とは異なり、各言語内での音長に関する規定に従う。全ての言語が音節をもっているが、音節とは異なるモーラをもつ言語ともたない言語がある。日本語学などでは、モーラを拍と呼ぶことが多い。また、日本語話者が日本語における音を数える際に、無意識に単位としていることが多くみられる。例えば、日本語定型詩の「七五調」や「五七調」、俳句の「五・七・五」、短歌の「五・七・五・七・七」などは、（しばしば無意識に「文字」などと言われることがあるが）実際にはこの拍を数えたものである。日本語の多くの方言においても同様である。日本語の仮名1文字が基本的に1拍である。ただし、捨て仮名（「ぁ」「ぃ」「ぅ」「ぇ」「ぉ」「ゃ」「ゅ」「ょ」「ゎ」といった小書きの仮名）は、その前の仮名と一体になって1拍である（たとえば「ちゃ」で1拍。拗音も参照）。一方、長音「ー」、促音「っ」、撥音「ん」は、独立して1拍に数えられる（これが「音

節」と異なる主な点である）。音節単位で見るなら、長音は長母音の後半部分を、促音は長子音の前半部分を切り取ったものであり、撥音は音節末鼻音や鼻母音をモーラとしたものといえる（鼻母音は基になる母音＋「ん」の2モーラになる）。これらは、「語頭に現れない」「単独で音節を形成しない」「お互いに連続することが稀である」などの性質をもち、二重母音の第二要素も含めて特殊拍（special mora）と呼称される。これらを除いて、単独で音節を形成する拍は自立拍（independent mora）と呼称される。表39にモーラの例を示す。

表39　　モーラの例

単語	音節区切り （音声学上の単位）	モーラ（拍）方言での区切り （いわゆる東京弁。 現代の俳句や短歌での 「七五調、五七調」の数え方）
さる（猿）	サ｜ル	サ｜ル
かっぱ（河童）	カッ｜パ	カ｜ッ｜パ
チョコレート	チョ｜コ｜レー｜ト	チョ｜コ｜レ｜ー｜ト
がっこうしんぶん（学校新聞）	ガッ｜コー｜シン｜ブン	ガ｜ッ｜コ｜ー｜シ｜ン｜ブ｜ン
がっきゅうしんぶん（学級新聞）	ガッ｜キュー｜シン｜ブン	ガ｜ッ｜キュ｜ー｜シ｜ン｜ブ｜ン
かんそく（観測）	カン｜ソ｜ク	カ｜ン｜ソ｜ク
かあさん（母さん）	カー｜サン	カ｜ー｜サ｜ン
にいさん（兄さん）	ニー｜サン	ニ｜ー｜サ｜ン

話速の自動推定のための特徴量としては、ビーウィズ社より提供されているコールセンターオペレーターの発話音声に対して音声認識より出力された発話テキストデータを使用する。この発話テキスト情報は漢字が含まれているため、まずはエクセルの関数でカナ文字を取得する。このカナ文字より、VBAでモーラ数を算出する関数を作成しモーラ数を取得し、それを発話継続時間で割り、単位時間あたりのモーラ数を算出する。そして、話速の自動推定の特徴量として以下を使用する。

・発話継続時間
・モーラ数
・モーラ数/発話継続時間（単位時間あたりのモーラ数）

自動推定の実行結果は以下である。機械学習手法としてはAdaboostを使用し、弱学習器として決定木との組み合わせで実施した。検証方法はk-分割交差検証法（k-Fold-CV）で実施した。その実験結果は表40である。

表40　　話速実験結果

速さ		推定結果		
		1	2	3
正解	1	7	20	13
	2	0	322	463
	3	0	163	1146

再現率	適合率	F値
0.175	1	0.298
0.41	0.638	0.499
0.875	0.707	0.782

正解率
69.1

全体的な正解率と評点１の再現率に課題が残る結果となった。これは、文章中に含まれる句読点に対するポーズ長が考慮できてないことが原因ではないかと考える。[22]によると、日本語音声の速さはしばしば、単位時間あたりに話されたモーラ数で算出される発話速度と、発話全体からポーズ時間を除いた時間長で算出される調音速度で表される。この単位はどちらも日本語音声ではモーラ数などを基準に算出されるもので、その音声が早く聞こえるのか遅く聞こえるのかという知覚的な側面とは必ずしも一致しない。音声言語の速さについて、音の高さの知覚的尺度であるメル、音の強さの聴覚的単位であるホンのような知覚量をあらわす単位は存在しない。しかしこれまで、日本語音声の速度感が何によって規定されるのか、もしくは影響されるのかについて、さまざまな視点から述べられてきた。当参考文献[22]によると

・ピッチ変動の大きい音声のほうが速く聞かれる可能性
・発話速度や調音速度が同じでもポーズ数が少ないほど「遅い」と知覚される
・発話速度の知覚に影響を与える要因として「ポーズ比」「ポーズ／秒 」「モーラ秒 」を挙げ、発話中のポーズの数や長さも重要

と述べられており、今回の実験ではポーズ長の考慮はできておらず、精度向上に向けて今後の課題であると考える。

3.4　本章のまとめ・考察

本章では声の大きさ、語頭、語尾、滑舌、抑揚、話速の評価項目に対する自動推定の方法とその実験結果について述べた。各節で説明した音響特徴量がそれぞれの評価項目の自動推定において一定程度有効であることがわかった。課題としては、全体的に評点１の再現率が低いことであり、これは評点クラスのデータ不均衡が影響している可能性がある。その改善案については第4章で述べる。また、語頭、語尾、話速の評価項目については音声データから時間方向の音素アライメントを取得し、より正確に各評価項目に対する特徴量を生成する必要がある。これは本論文では実験できていないため、今後の課題であると考える。

第４章　不均衡対策について

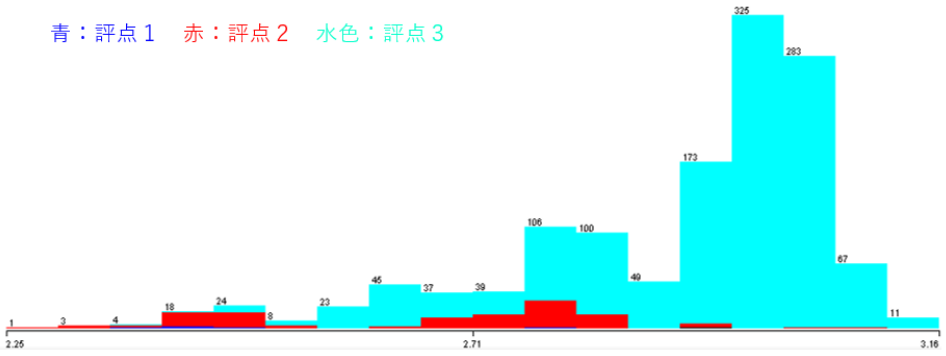
4.1　回帰分析の活用提案

第3章では、各評価項目に対する自動推定の実験と結果を記載した。どの評価項目でも課題であるが、評点１，２，３の各データ件数が不均衡であり、全体的にそれが原因で精度向上のボトルネックになっている可能性がある。そのため、本章では不均衡データに対する精度改善策として、①回帰分析の活用提案と、②コスト考慮型学習の二つについて提案法として述べる。

まずは、本節では①の回帰分析の活用提案について述べる。実験の対象は評価項目１「声が大きすぎないか」とした。データの不均衡を解消するため、評点２の一部のデータを評点１とみなし件数を増やす案を考えた。但し、評定２の内でどの対象を評点１へ付け替えればモデルへの影響が少ないかを考慮する必要がある。そこで、評点１，２，３を分類ラベルとしてではなく、評点１～３の間の連続量ととらえ、回帰分析によりデータの評点をソフトに判別することを考える。回帰により、評点1.2、評点2.7等、1～３の間に推論される対象が把握できるため、評点２の内、回帰により評点が1に近しい値に推論された対象データを評点1へ付け替え、評点１の件数増加を試みる。

図25が回帰分析を実施した結果である。元ラベルの評点と、回帰で得られた評点にある程度の相関はみられるが、評点１，２，３を明確に区分できるほどの結果ではなかった。全体的に高い評点で推論されており、評点３の件数が多いことによると思われる。

図25　　評点の回帰分析



そこで、次のような工夫を考えてみた。評点３を１００点満点とし、各ラベルの発生件数割合を加味して相対的な評価得点を設定する。相対評価得点の設定内容を表41に示す。

相対評価得点＝評価得点×相対評価得点（件数が少ないほど得点は減少する）

上記の相対評価得点を目的変数として、同様に回帰分析してみる。

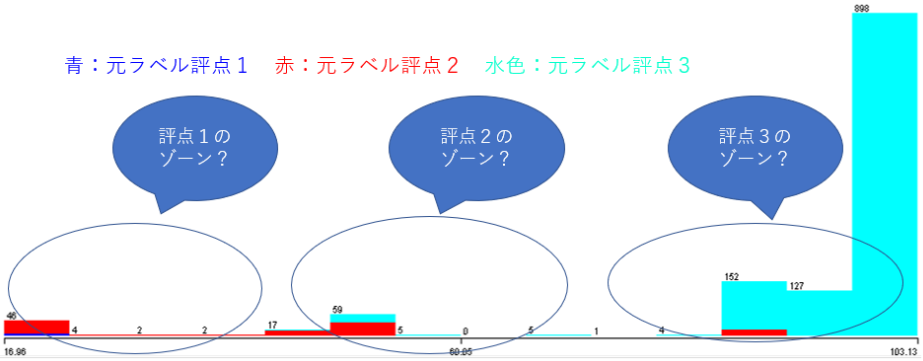
表41　不均衡を考慮した相対評価得点

これを目的変数

元ラベル	評価得点	相対評価得点	相対件数割合
3	100	100	1
2	70	7	0.1
1	30	1.5	0.05

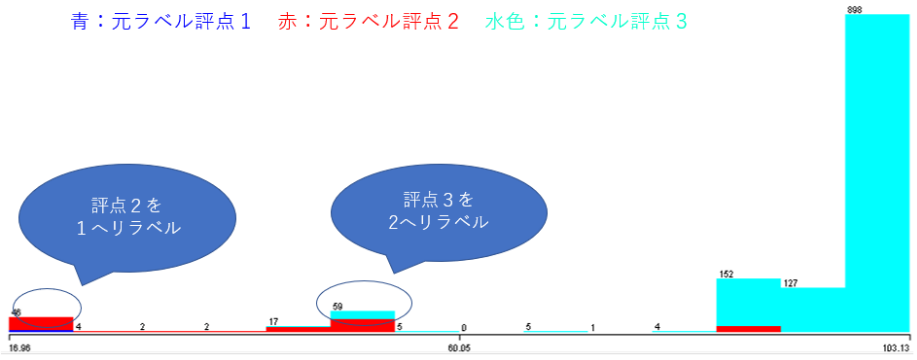
その回帰分析の結果が図26である。元評点１，２，３のまま回帰した結果に比べ、相対評価得点を目的変数として回帰した場合は、その分布が元ラベルの評点に応じてより明確になった。

図26　相対評価得点による回帰分析



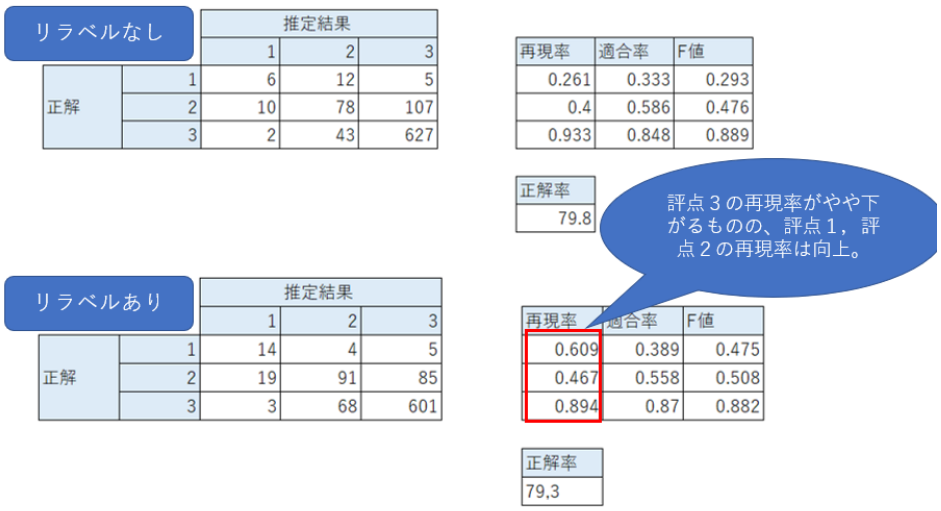
その結果を活用し、図27のように得点の一番低いゾーンにある元ラベル２を１へ、2番目に低い元ラベル３を２へリラベルし、このラベリング情報で、再度１，２，３の分類問題として学習し、テストしてみた。リラベルした件数は78件であり、全体件数1311件に対して5.9%をリラベルしたことになる。人間の認識誤り率が5%であることを踏まえると、リラベルした対象はNoisyとみなし無理のない件数であると考える。

図27 相対評価得点による回帰分析（リラベル対象）



そのテスト結果が表42である。機械学習手法としてはAdaboostを使用し、弱学習器として決定木との組み合わせで実施した。検証方法はk-分割交差検証法（k-Fold-CV）で実施した。上段がリラベルなしでモデルを構築しテストした結果、下段がリラベルありでモデルを構築しテストした結果である。結果としては、評点３の再現率がやや下がるものの、評点１，評点２の再現率が向上することになり、全体的にバランスのよいモデルが作れたことになる。よって、回帰による得点分布の取得と、それに応じたリラベルはモデル全体の精度向上に効果があることがわかった。

表42 リラベル有無による実験比較



このように、目的変数に相対評価得点を導入し、分類問題から回帰問題とすることで、一部の目的変数をリラベルし不均衡を解消することで、評点１の再現率が向上することがわかった。相対評価得点の導入は目的変数を非線形空間へ落としこむことを意味している。そこで、当初の目的変数を線形空間のままで実施したケースと、非線形空間で実施したケースで比較してみることにした。ここでは、リラベルはせずに目的変数の線形空間と非線形空間の回帰問題としての比較による効果を確認してみた。比較結果は表43である。対象は同じく評価項目１「声が大きすぎないか」で実施した。評点１，２，３の分類問題として実施した結果と、それを回帰問題として分類した結果、そして相対評価得点という非線形な目的変数での回帰問題として分類した結果である。結果としては当初の分類問題と比べ大きな精度改善の効果はみられなかった。

表43　　分類問題と回帰問題の比較結果

次に、リラベルによる効果を確認するため、回帰の線形問題において一部データをリラベルした場合と、回帰の非線形問題において一部データをリラベルした場合とで当初の分類問題に対して精度向上の効果があるかを実験してみた。リラベルした対象は、当節の冒頭で実験した時と同様に、得られた連続回帰値から評点が低い元ラベル２のデータを１へ、元ラベル３のデータを２へリラベルして実験した。その結果が表44である。リラベルにより評点１の適合率や評点２の再現率や適合率が改善し、また全体の正解率も約2%向上している。

このように、分類問題を回帰問題へ変換するだけでは精度改善はみられなかったが、回帰問題により一部データをリラベルすることは精度改善に効果があることがわかった。これは、自己適応型学習[25]と同様の効果があるためだと考えられる。自己適応型学習とは、適度に訓練された（過適合されていない）ニューラルネットワークの出力は真の確率ベクトル（各訓練ラベルに対する帰属度）を近似している仮定し、訓練ラベルをある定式で順次修正しながら学習するという手法である。今回の提案法では、分類問題における訓練ラベルを、いったん回帰問題と考えることで連続回帰値を取得し、それをソフトラベルと考え、値の近しいソフトラベルを同じ評点のカテゴリと見なすことで、訓練ラベルを一部修正した上でモデルを学習するという仕組みにより、データ数の少ない評点１の再現率向上を実現させている。

表44　　分類問題と回帰問題のリラベル比較結果

分類問題

正解＼推定結果	1	2	3
1	10	26	6
2	16	140	156
3	2	112	1643

再現率	適合率	F値
0.238	0.333	0.278
0.461	0.519	0.489
0.935	0.91	0.923

正解率
84.9

線形リラベル

正解＼推定結果	1	2	3
1	10	29	3
2	2	203	118
3	1	88	1658

再現率	適合率	F値
0.192	0.769	0.308
0.628	0.615	0.622
0.949	0.932	0.94

正解率
88.1

非線形リラベル

正解＼推定結果	1	2	3
1	18	19	5
2	0	195	128
3	0	86	1661

再現率	適合率	F値
0.346	1	0.514
0.603	0.629	0.616
0.951	0.926	0.938

正解率
88.1

4.2　コスト考慮型学習について

本節では、もう一つのデータ不均衡対策としてコスト考慮型学習を提案する。今回の分類問題では、評点１，２，３を目的変数のラベルとして学習してきたが、単純に分類する場合、各クラスを等価なものとして扱うことになってしまう。例えば、評点１のデータが評点２に誤分類されるケースと、評点１のデータが評点３に誤分類されるケースでの損失関数の計算は同じである。しかし、評点１のデータは印象評価において「相手に悪い印象を与えるケース」として、評点３である「問題ないケース」への誤分類は、評点２への誤分類に対して極力少なくしたい。このように、評点の各クラスを等価なものとして扱うのではなく、評点間に差を設けることで精度向上する策を考えてみた。

そこで、冒頭で話したコスト考慮型学習で実験してみることにした。コスト考慮型学習（Cost-Sensitive Learning）とはコストを定義して、そのコストを用いて仮説を得る手法のことである。コストとして様々なものが提案されており、例えば誤分類コスト、データ取得コスト、計算コストなどが提案されており、目的に応じて使い分けたり、あるいは自ら設計することもできる。 Cost-Sensitive Learningは大きく分けると2つのグループにカテゴライズできる[26]：

Direct approaches： 学習アルゴリズムにコストを取り入れるアプローチ。学習アルゴ

リズムの修正を伴うため、学習器依存。

Meta-learning approaches： 学習データや仮説の出力値を修正するアプローチ。学習の前処理と後処理に相当し、学習アルゴリズムの修正を伴わないので、任意の学習器に適用可能。

今回は後者のMeta-learning approachesで考えてみることにした。2章で紹介したデータマイニングツールWekaにCostSensitiveClassifierというメタ学習機能がありそれを利用する。CostSensitiveClassifierはメタ学習なので、弱学習器としてSVMや決定木等広く組み合わせ可能であり、コスト行列の設定において、少数派クラスに対する誤分類コストを変更することで不均衡に有効であると考えた。今回は、弱学習器として決定木を設定し、また、コスト行列として以下の様に少数派クラスである評点１の誤分類コストを他評点に比べ高く設定して実験してみた。検証方法はk-分割交差検証法（k-Fold-CV）で実施した。コスト行列の設定内容を表45に示す。

表45　　コスト行列の設定：通常とコスト操作

通常コスト		予測クラス		
		1	2	3
真のクラス	1	0	1	1
	2	1	0	1
	3	1	1	0

コスト操作		予測クラス		
		1	2	3
真のクラス	1	0	2	3
	2	1	0	1
	3	1	1	0

その実験結果が表46である。評価項目８語尾上がりのケースで実験した。結果としては評点１の再現率を改善することができている。

表46　　コスト操作有無による実験結果

通常コスト		推定結果		
		1	2	3
正解	1	3	4	2
	2	2	22	14
	3	2	8	235

再現率	適合率	F値
0.33	0.429	0.375
0.579	0.647	0.611
0.959	0.883	0.948

正解率
89

コスト操作		推定結果		
		1	2	3
正解	1	6	2	1
	2	4	22	12
	3	4	10	231

再現率	適合率	F値
0.667	0.429	0.522
0.579	0.647	0.611
0.943	0.947	0.945

正解率
88.6

また、評点１が評点２へ誤分類されるよりも、評点３へ誤分類されるときのコストを高くすることによる効果を確認するため、評点２と評点３への誤分類コストを同じにしたケースと、差をつけたケースで比較してみた。その実験結果が表47である。評点１と評点３でコストに差を設けた場合、評点１が評点３に誤分類されるケースが減り、差を設けない場合と比べ再現率が高いことがわかる。このように、多値分類問題において各クラス間を等価に扱うのではなく、コスト計算等で差を設ける手法は不均衡問題で有効であることがわかった。今回行った方法以外では、各クラス間に距離の概念を導入し、クラス間の誤分類における計算で差を設ける順序回帰問題[27]等の方法もある。また、今回はコスト行列を手動で設定して実験したが、コスト行列自体の計算を自動で行う方法も提案されている[28]。

表47　　コスト評点間差異の有無による実験結果

コスト考慮１

		推定結果		
		1	2	3
正解	1	14	3	6
	2	12	94	89
	3	2	41	629

再現率	適合率	F値
0.609	0.5	0.549
0.482	0.681	0.482
0.936	0.869	0.901

正解率
82.8

		予測クラス		
		1	2	3
真のクラス	1	0	2	2
	2	1	0	1
	3	1	1	0

コスト考慮２

		推定結果		
		1	2	3
正解	1	18	2	3
	2	14	93	88
	3	2	41	629

再現率	適合率	F値
0.726	0.543	0.655
0.477	0.689	0.564
0.936	0.874	0.904

正解率
83.2

		予測クラス		
		1	2	3
真のクラス	1	0	2	3
	2	1	0	1
	3	1	1	0

21

4.3　本章のまとめと考察

本章では、第3章での各評価項目に対する実験結果における課題であるデータ不均衡問題の解消に対して２つの提案法について述べた。一つ目は分類問題を回帰問題と考え、回帰値に応じて一部のデータをリラベルすることで、小データ群である評点１の再現率の改善に有効であることがわかった。２つ目はコスト考慮型学習で各評点間の誤分類におけるコストに差を設けることで、評点１の再現率の改善、及び評点１が評点３へ誤分類される件数を減らすことに有効であることがわかった。

第５章 結論

本研究ではコールセンターにおけるオペレーターの応対品質評価のうち自動化されていない対象の自動化を目標として，音響解析型の技術を応用してその実現可能性を検討した。

第2章では自動化対象である評価項目1〜18（声の大きさ、語尾、抑揚、滑舌、話速）と重要な関係のあるパラ言語についてその概要を述べた。また、本研究で使用する音響特徴量について述べた。特にopenSMILE ツールキット（音声信号から特徴量を抽出できるオープンソースのツールキット）と呼ばれるツールにより取得できるIS09,IS10特徴量を利用するためそれについて述べた。最後に、データマイニングツールWekaを使用しているため、その概要を述べた。後続の第3章では、これら音響特徴量や実験環境を用いて印象評価の自動推定実験をおこなっている。

第3章では声の大きさ、語頭、語尾、滑舌、抑揚、話速の評価項目に対する自動推定の方法とその実験結果について述べた。各節で説明した音響特徴量がそれぞれの評価項目の自動推定において一定程度有効であることがわかった。課題としては、全体的に評点１の再現率が低いことであり、これは評点クラスのデータ不均衡が影響している可能性がある。その改善案については第4章で述べる。また、語頭、語尾、話速の評価項目については音声データから時間方向の音素アライメントを取得し、より正確に各評価項目に対する特徴量を生成する必要がある。これは本論文では実験できていないため、今後の課題であると考える。

第4章では、第3章での各評価項目に対する実験結果における課題であるデータ不均衡問題の解消に対して２つの提案法について述べた。一つ目は分類問題を回帰問題と考え、回帰値に応じて一部のデータをリラベルすることで、小データ群である評点１の再現率の改善に有効であることがわかった。２つ目はコスト考慮型学習で各評点間の誤分類におけるコストに差を設けることで、評点１の再現率の改善、及び評点１が評点３へ誤分類される件数を減らすことに有効であることがわかった。

謝辞

本研究を進めるにあたり，多くの方にご協力を賜りました。ここに，心より感謝の意を表します。特に，滋賀大学大学院データサイエンス研究科 市川治教授には，指導教員として多大なご指導とご支援を頂きました。研究を進めるための貴重なアドバイスを頂き，思うような進捗が出せなかった時にも温かい励ましの言葉を常に掛けて頂きました。心から感謝申し上げます。また，本研究を進める為に不可欠なコールセンターの音声データをご提供頂き，その全てに機械学習用の詳細なラベルデータを付与して下さり，現場における品質評価の実際や評価のポイントをご教示下さりましたビーウィズ株式会社のご担当者の皆様に深く感謝申し上げます。深く感謝申し上げます。

参考文献

[1] 富士通株式会社, “応対自動評価システム”, URL:https://www.fujitsu.com/jp/services/application-services/enterprise-applications/crm/voicetracking/qualitymanager/

[2] 株式会社日立情報通信エンジニアリング, “音声分析サービス for コンタクトセンター”, URL:https://www.hitachi-ite.co.jp/products/voice_analysis/index.html

[3] 鈴木基之, “音声に含まれる感情の認識”, 日本音響学会誌 71 巻 9 号(2015), pp.484-489, 2015.

[4] 有本泰子 *et al*, “「怒り」の発話を対象とした話者の感情の程度推定法”, 自然言語処理 Vol.14 No.3, pp147-163, 2007.

[5] 岡田敦志 *et al*, “表情・音響情報・テキスト情報からのリアルタイム感情推定システム”, *The 31st Annual Conference of the Japanese Society for Artificial Intelligence* , 2017.

[6] 森大毅, “感情音声の研究を始める人のための音声コーパス入門”, 日本音響学会 2019 年春季研究発表会スペシャルセッション[音声A／音声B], URL:https://speakerdeck.com/hirokisince1998/gan-qing-yin-sheng-falseyan-jiu-woshi-meruren-falsetamefalseyin-sheng-kopasuru-men

[7] 株式会社AGI, “定量精神分析研究の動向”, URL: https://www.agi-web.co.jp/technology/trend.html

[8] 池本真知子 *et al*, “感情判別における声質の影響”, 感情心理学研究 2009 年 第 16 巻 第 3 号, pp.209-219, 2009.

[9] 森大毅, “音声からの感情・態度の理解”, 電気情報通信学会誌 Vol.101 No.9, 2018.

[10] 有本泰子 *et al*, “音声チャットを利用したオンラインゲーム感情音声コーパス”, 日本音響学会講演論文集 2013 年 9 月, pp.385-388, 2013.

[11] Emika Takeishi, "Construction and Analysis of Phonetically and Prosodically Balanced Emotional Speech Database", *Proceedings of Oriental COCOSDA*, pp.16-21, 2016.

[12] 日本声優統計学会, “声優統計コーパス”, URL:https://voice-statistics.github.io/

[13] "感情評定値付きオンラインゲーム音声チャットコーパス（OGVC）", URL: https://sites.google.com/site/ogcorpus/

[14] “UUDB 宇都宮大学 パラ言語情報研究向け音声対話データベース”, URL: http://uudb.speech-lab.org/index.html

[15] 西川仁, 佐藤智和, 市川治, 清水昌平, “テキスト・画像・音声データ分析”, pp.139-178, 講談社, 2020.

[16] B.Schuller *et al*, "The Interspeech 2009 emotion challenge", *Proc.INTERSPEECH*, pp.312-315, 2009.

[17] B.Schuller *et al*., “The INTERSPEECH 2010 Paralinguistic Challenge”, *Proc. Interspeech*, pp. 2794-2797, 2010

[18]千吉良好紀 *et al*, “漸次的な感情認識法における excitation pattern と基本周波数の利用法の検討”, 日本音響学会講演論文集 2014 年 9 月, pp.387-388, 2014.

[19]竹部真晃 *et al*, “音声感情認識における声門特性に基づく特徴量の検討”, 日本音響学会講演論文集 2015 年 9 月, pp.113-116, 2015.

[20]羽田優花 *et al*, “日本語感情音声コーパス JTES を対象とした感情認識の基礎検討”, 情報処理学会東北支部研究報告 Vol.2019 No.A3-1, 2019.

[21]武石笑歌 *et al*, “感情音声データベース構築に向けた音韻・韻律バランス感情音声の収録と分析”, 日本音響学会講演論文集, pp.355-358, 2016.

[22]時間構造分割特徴量に基づく感情発声の自動分類」, 原, 伊藤, 2011年, 日本音響学会春季研究発表会）

[23]速度変化をともなう音声の速度感とその規定要因 丸島 歩

[24]発話音声の聞き取りやすさ向上のための音声特徴量解析 佐賀 圭真 井村 誠孝 エンタテインメントコンピューティングシンポジウム (EC2019)」2019 年 9 月

[25]Self-Adaptive Training: beyond Empirical Risk Minimization Part of Advances in Neural Information Processing Systems 33 (NeurIPS 2020)

[26]A Cost Sensitive Technique for Ordinal Classification Problems : Sotiris B. KotsiantisPanagiotis E. Pintelas

[27]Cost Sensitive Learning of Deep Feature Representations from Imbalanced Data : Salman H. Khan, Munawar Hayat, Mohammed Bennamoun, Ferdous Sohel, Roberto Togneri

[28]Ordinal Regression Methods: Survey and Experimental Study : July 2015IEEE Transactions on Knowledge and Data Engineering 28(1)

[29]コールセンターの応対音声品質の自動評価に向けた要素技術の研究 高山 和明