アナウンス技術の練習支援システムに向けた基礎検討

山田　真那子[1,a)]　塚田　浩二[1]

概要：
本研究では，音声情報処理技術を用いたアナウンスの練習支援システムを提案する．アナウンスとは一般的にテレビやラジオなどでアナウンサーが情報を分かりやすく伝えるための技術である．本研究では，高校や大学で行われる競技アナウンスの採点基準を参考にして，アナウンス技術の評価指標を考察する．さらに，アナウンス初心者／経験者の音声データを収集し，評価指標に基づいて実装したシステムを用いて，アナウンス技術の定量的な分析を試みる．こうした分析を通して，アナウンス技術の練習支援システムに向けた基礎検討を行う．

Basic study toward a practice support system for announcing techniques

MINAKO YAMADA[1,a)]　KOJI TSUKADA[1]

1. はじめに

アナウンスは，聞いている人に情報を分かりやすく伝えるための技術であり，テレビやラジオのアナウンサーに必須とされるだけでなく，高校や大学でもアナウンス技術を競う大会がある [1][2]．アナウンスの練習では，熟練者のアナウンス音声を聞いて真似したり，熟練者からアドバイスを受けることが一般的である．しかし，熟練者のアドバイスは抽象的な表現を使った感覚的なものが多く，初心者が正確に理解することは難しい．例えば，初心者に対するアドバイスとして多いのが「文にうねりがあったから，しっかり下そう」というものがある．これは「文の音を一定に下げていかなければいけないのに，途中で音が上がってしまっていたので音を下げよう」という意味である．途中で音が上がるという意味の「うねり」や，文の音を下げながら読む「下し」などの専門的な表現が，初心者にとっては難しい．また，文章内のどの部分でどの程度音を下せていないのかも理解しづらい．そのため，初心者が感覚を掴めるようになるまで膨大な時間をかけて，熟練者と感覚をすり合わせるような繰り返しの練習が必要になる．更に，一人で練習をすると無意識に悪い癖がついてしまうことがあり，こまめに熟練者に練習を聴いてもらう環境が必要になる．

本研究では，アナウンス技術を要素毎に分類・定量化し，学習者のアナウンスを分析／フィードバックすることで，一人でのアナウンス技術習得を支援するシステムの構築を目標とする．本論文では支援システムの実現にむけた基礎検討として，アナウンス技術の定量的な評価指標を提案・検討する．

2. 関連研究

本章では，本研究に関連する研究事例を，「歌唱や発話の支援」,「音声素材の音響的分析」の 2 つの観点から説明する．

2.1 歌唱や発話の支援

歌唱支援に関する事例として，中野ら [3] は，ユーザの歌唱力向上を支援するインタフェース Mirusinger を提案している．既存の楽曲のボーカルパートの音高（F0）を分析・可視化し，ユーザ自身の歌声と比較表示することができる．また，羽賀ら [4] は，歌唱の自主練習における学習

[1] 公立はこだて未来大学
Future University Hakodate
[a)] g2124053@fun.ac.jp

効率の向上を目的とし，習熟度に関係する音響特徴量を可視化することで，学習者の歌唱データを評価した．プレゼンテーションに関する事例として，栗原ら [5] は，音声認識／画像認識を用いて，プレゼンテーション中の発表者の様子を分析し，リアルタイムにフィードバックするシステムを提案している．具体的な指標としては，話速度・声の抑揚・聴衆とのアイコンタクト等を用いている．また，趙ら [6] は KINECT センサを利用し，身体表現の「アイコンタクト」と「体の向き」，音声表現の「声の大きさ」を支援している．例えば，発表練習における良くない状態の時間が累積した場合，イヤフォンから音声で「前へ」，「大きい声」などと再生される．

さらに，声優の発話支援として，滑舌練習を行える「声優滑舌アプリ」[7] や，アフレコの練習を行うことが出来る「SAY-U」[8] などがある．一方，これらは音声分析等の知的機能は搭載されていない．また，声優では演技が重視されるのに対して，アナウンスは自然な発声で情報を正確に伝えることが求められるため，適した支援手法は変わってくると考える．

2.2 音声素材の音響的分析

王 [9] は，ナレーション（映像の説明音声），アナウンス（原稿の忠実な読み上げ），声優（声による演技）の 3 種類の音声データを対象として，音響的分析を行っている．例えば，声優はアナウンサーより母音の発音が長い．しかし，文末の句点による静止時間については三者とも長く，2 秒ほどとっていると述べている．アナウンスの初心者にとって間（無音時間）をしっかりとるのが難しいとされるが，アナウンス以外のナレーション／声優も，熟練者はしっかりと間を取っていることが分かる．

2.3 本研究の特徴

プレゼンテーションや歌唱支援の支援システムは多く開発されているが，アナウンスに焦点を当てた支援システムは調べた中では開発されていない．本研究では最終的に，アナウンスに特化した支援システムの開発を目指す．

3. 提案

本研究では，アナウンス技術を要素毎に分類・定量化し，学習者のアナウンスを分析／フィードバックすることで，技術習得を支援するシステムの提案を目指す．本稿ではまず，「熟練度の異なるアナウンス音声のデータ収集と評価指標の考案」，「音声認識と音響分析を用いた評価技術の設計」について述べる．

3.1 データ収集と評価指標の考案

大学内の放送サークルのメンバーの協力や，オンライン上で公開されているアナウンス熟練者（大会成績上位者）の音声データを活用して進める．

評価指標は，NHK 杯全国高校放送コンテスト [1] や NHK 全国大学放送コンテスト [2] におけるアナウンス部門の評価基準を参考に検討する．表 1 に，既存の評価基準を整理したものを示す．次に，各評価指標について詳しく説明する．

表 1 放送コンテストの審査基準

審査基準	補足説明
発声	音量の一定性，語尾の発声の仕方 （伸びていないか，促音が入っていないか）
発音	滑舌（一音一音の明瞭さ）
アクセント	単語の音の高低の正しさ
イントネーション	一文の音が適度に下がっていくか
テンポ	文章を読む速度
ポーズ	句読点での間（無音時間の長さ）
言葉の立て方	重要な言葉を強調できているか （強調したい言葉の速度を遅くする， 音を高くする，直前に間を入れる）

3.1.1 発声

発声には，音量の一定性と語尾の発声の仕方の要素がある．音量の一定性は，声が大きすぎたり，小さすぎないように発声しているかという基準である．アナウンスでは常に一定の音量での発声が必要とされる．語尾の発声の仕方は，例えばよくある語尾の「〜します。」という表現があった場合に，「しますぅ。」という風に最後の一音の母音が残ってしまう場合や，反対に「しますっ。」といった風に最後の一音を勢いよく発声してしまい，語尾に促音が入っているように聞こえてはいけないということである．このような発声を避けるために，指導では「ペンをゆっくり机に置くようなイメージで発声しよう。」ということがある．

3.1.2 発音

発音とは，滑舌のことである．放送コンテストなどでの滑舌は，日常生活の会話のように聞き取れればいいということではなく，一音一音に子音が混じっていないか，明瞭に発音できているかを注意深く評価される．

3.1.3 アクセント・イントネーション

アクセントとイントネーションはどちらも音の高低に関する項目である．

アクセントは個々の語について，社会的慣習として決まっている相対的な音の高低で表現される．例えば，「橋」・「箸」・「端」という全てが「はし」と読まれる単語でも，音の高低によって異なる意味を持つ．アクセントは通常，日本語発音アクセント新辞典 [10] などのアクセント辞典によって単語ごとに指定されており，音の下がる位置や鼻濁音，無声音の指示が示されている．また，アクセントには第一アクセント，第二アクセントなど一つの単語による複数のアクセント表記があるものもある．これらはどちらで

読んでもいいとされる．

イントネーションは文の音の高低が文意に影響を与える要素である．アナウンスでは，通常，文頭の音が高く，文末に向かって音の高さが低くなる傾向が見られる．また，文章の係り受けの関係を示すために，文の途中で音を上げる技法も利用される．目立たせる言葉の前では音を高くするなど，様々な場面で音の高低を変化させる必要がある．

3.1.4 テンポ

審査基準では文章を読む速度の具体的な指定はないが，聞き取りやすい速度で読むことが必要とされている．

3.1.5 ポーズ

アナウンスでは，様々な箇所で間（無音時間）をおくことで，聞き取りやすく，情報を聴き手に分かりやすく伝えるようにしている．基本的には句読点などで間をおくが，こちらも放送コンテストなどでの具体的な箇所・秒数は指定されていない．

3.1.6 言葉の立て方

言葉の立て方は，言葉を強調する際に使う技法である．“遅く読む”・“高く読む”・“前に間を入れて読む” の 3 つで言葉を強調出来ると考えられる．

アナウンスでは，文章を分かりやすく伝えるために，情報の優先度をつけて読む必要がある．そのため，固有表現（固有名詞や，数詞など）など原稿の重要な部分を上記の方法で強調する．特にアナウンスでは初出の情報が多く，登場する人物名や，その出来事に関連する日時を分かりやすく，印象に残るように伝えなければいけない．また，強調の表現方法として上記 3 つが挙げられるが，場合に応じて，どれか一つを使う場合もあれば，3 つを組み合わせて強調する場合もある．

“遅く読む” というのは比較的使用しやすい技法で，シンプルに強調したい単語をゆっくり読むことである．“高く読む” は，単語全体の音を高くするというよりは，イントネーションの箇所で説明した通り，本来文章を読んでいくにつれて音が下がっていくのを，強調したい単語の 1 音目で少しだけ音を上げて，また音を下げながら文章を読んでいくことである．つまり，強調したい単語のみを高くするのではなく，強調したい単語で音が上がり，また自然に音を下げていくイメージである．“前に間を入れて読む” というのは，前述したポーズよりは短めの無音時間を入れることで言葉を強調する技法である．

3.2 システム構成

本研究では，アナウンス技術を要素毎に分類・定量化し，学習者のアナウンスを分析／フィードバックすることで，技術習得を支援するシステムの提案を目指す．

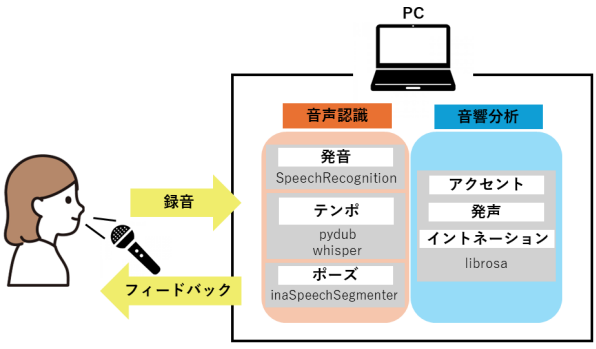

図 1 システム構成図

開発環境としては，Python で音声認識と音響分析を利用する．

音声認識に関しては，発音は SpeechRecognition ライブラリを利用し，アナウンスの音声データを入力して文字起こしを行い，原稿データと比較して差異の分析を行う．テンポは pydub ライブラリを利用し，音声データの長さと原稿データの字数から音声全体の話速度を求める．今後 whisper ライブラリを利用し音声データから各文ごとのタイムスタンプを取得し，詳細な話速度を分析する．ポーズは inaSpeechSegmenter ライブラリを利用し音声データから無音時間を検出する．

音響分析に関しては librosa ライブラリを利用する．アクセント，イントネーションについては音声データからピッチ (f0) の抽出を行い，正しいアクセントや本研究で設定したイントネーションとの比較を行う．発声は音声データの音圧 (dB) を出力して，標準偏差等を計算する．

4. 実装

本章では，アナウンスの音声データの収集方法や，評価指標の考案，及び評価技術の実装の詳細について紹介する．

4.1 データ収集と前処理

大学内の放送サークルのアナウンス初学者・経験者のアナウンスをマイクで録音し，アナウンス音声を収集する．現状では，アナウンス初学者 3 人，アナウンス経験者・熟練者 7 人のデータを収集している．詳細は 5 章の評価実験で述べる．また，熟練者のアナウンス音声については過去の大会のものが CD や YouTube 上に公開されているため，今後はそれらも利用する予定である．

大会などの音声データは，名前，番号，タイトル等の審査に関係しない読み上げがあるため，前処理としてこれらを削除する．将来的には自動化を進めていくが，現段階では，音声編集ソフトを用いて手動でトリミングしている．

4.2 評価指標と評価技術

表 1 で整理した審査基準に基づき，具体的な評価指標を表 2 のように設定する．評価指標に合わせて，評価技術の開発を行う．評価技術で利用する Python ライブラリは表 2 の通りである．

表 2 評価指標と評価技術

審査基準	本研究における評価指標	評価技術
発声	・音量が一定になっているか	librosa
発音	・音声認識によって文字起こしされた原稿と，本来のアナウンス原稿に差異が無いか	SpeechRecognition
アクセント	・単語の音の高低があっているか	librosa
イントネーション	・読点で区切った場合に，音が下がっていっているか	librosa
テンポ	・1 分間あたり 300 字程度で読めているか	pydub
ポーズ	・句点での間は 2 秒，読点での間は 1 秒程度とれているか	inaSpeechSegmenter
言葉の立て方	・固有表現を読む速度は遅くなっているか ・固有表現の音は高くなっているか ・固有表現の前に間を入れているか	(未対応)

4.2.1 発声

本研究では，発声の要素の中から，音量の一定性に着目する．アナウンスは感情表現を激しくしてはいけないとされる．そのため，言葉を強調したい場合にも音量を大きくする表現方法は望ましくない．さらに，初学者の特徴として，文頭の音を勢いよく発声してしまうため，文章の一音目が強くなってしまいがちである．また，アナウンス原稿の一文を読み切るのに，息が続かない・体力がないなどの要因から文末の音量が小さくなる傾向にある．これらの初学者特有の課題を解決するためにも，音量の一定性を発声の評価指標として設定した．

Python のライブラリ librosa を利用し，音量の標準偏差を求めることで，音量の一定性を分析する．アナウンスは音量が一定であるべきとされるため，音量の変動の程度が小さいほど評価が高くなる．また，録音環境による影響などを観察するために，最小値・最大値・平均値も共に求める．

4.2.2 発音

一音一音が明瞭に発音できているかを評価するため，音声認識による文字起こしされた原稿と本来のアナウンス原稿の差異を分析する．一音一音の明瞭さはいわゆる「滑舌」といい，この滑舌がよくないと間違った情報が伝わってしまう可能性があるため，審査でも重要視される要素である．

Python のライブラリ SpeechRecognition を利用して，アナウンス音声の文字起こしを行う．学習者が読む原稿データと文字起こしデータの差異を検証し，発音の明瞭さを評価する．発音が明瞭である程，文字起こしの認識精度が高くなり，差異は少ないと考えられる．なお，書き起こしたものが同音異義語になってしまう場合があるが，アクセントが同じ単語の場合は漢字変換の問題の為，差異としない．

4.2.3 アクセント

アクセントは，3.1 節でも述べた通り，決められた単語の音の高低などを正しく発音できているかを評価する．アナウンスは音声のみで情報を伝えるため，アクセントが適切でないと間違った情報が聞き手に伝わったり，違和感を与えて聞き逃しを引き起こす可能性がある．

Python のライブラリ librosa を利用し，周波数分析を用いた音の高低について分析を行う．1 分半のアナウンス音声のピッチ (f0) を可視化した例を図 2 に示す．現時点では，アクセントは単語ごとに NHK 日本語発音アクセント新辞典を用いて，手動で指定している．今後は，日本語音韻データベース[*1] 等を活用して，自動的に認識する実装を進めていきたい．

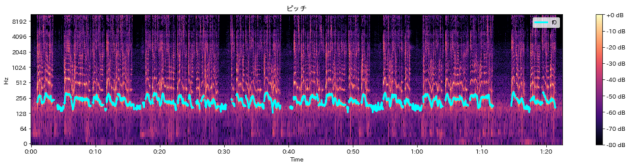

図 2 ピッチ抽出したアナウンス音声の例

4.2.4 イントネーション

イントネーションは文章の音が下がっていくかを評価する．アナウンスでは，文頭の音が高く，文末にいくにつれ徐々に音を下げていくのが基本とされている．適切でない箇所で音が高くなってしまうことを「うねり」といい，このうねりが発生すると聞きづらさに繋がってしまう．

また，適切な箇所で音を高くすることを「上げなおし」といい，これは文章の係り受けの関係を分かりやすくするためのもので，文の途中であっても文章の意味を分かりやすくするために音を高くする場合がある．主語と述語の関係を明瞭にするためや，並列の関係を示す場合に音の高さを調整することもあるので必ずしも常に音を下げるわけではない．しかし，文末に注目すると，文頭より必ず音が下がっている必要がある．

Python のライブラリ librosa を利用し，周波数分析を用いた音の高低について分析を行う．現時点では，目視で文末の変化を確認しているが，今後前述した SpeechRecognition 等を用いて句点ごとに文章を区切り音が下がっているか自動で確認する実装を行う．

4.2.5 テンポ

文章を読む速度を評価する．元 NHK アナウンサーであり，長崎大学でスピーチ研究を行っている矢野 [11] らは，文章を読んだときに聴き手が聞き取りやすい速度は，1 分間あたり 300 字程度としているため，これを目安に分析する．

原稿の文字数と音声データの長さから，1 分当たりに読む文字数を計算する．原稿には「」などの記号が含まれるため，発音しない文字は Python の re モジュール等で削

[*1] https://www.cjk.org/language/ja/data/japanese/nlp/japanese-phonetic-database/

除する．音声ファイルの再生時間は，pydub ライブラリの AudioSegment を利用し取得する．

更に，原稿全体の読むスピードだけでなく，一つの原稿内でスピードがどの程度変化も調べるため，文ごとのスピードを検知する機能の開発も進めている．Whisper ライブラリを使用し，タイムスタンプごとに含まれる文字数を計算する．

4.2.6 ポーズ

適切な間をとれているかを評価する．間は句読点でとることが多い．間の長さは秒数で一意に決めるものではなく，文章の意味を考えて，適切な長さにする必要がある．

本研究では，2.2 節で紹介した王 [7] の研究や著者の経験[*2]を基に，目安として句点で 2 秒，読点で 1 秒程度の間をとることを当面の方針とする．

句読点やその他の表現における無音時間の長さを分析する．目安として 4.2 節で句点での間は 2 秒，読点での間は 1 秒程度と設定している．これらの要素を考慮しながら，ライブラリ inaSpeechSegmenter を利用して無音時間を検知し，原稿データの句読点と対応した位置で適切に間を入れているかを評価する．今後は自動で無音時間と照らし合わせて句読点を自動分類していく．

4.2.7 言葉の立て方

言葉の立て方については，音響分析と音声認識だけでなく，自然言語処理などの技術も必要となるため，本論文では対象とせず，今後の開発課題とする．

5. 評価実験

この章では，評価技術に対する基礎的な性能評価について述べる．

5.1 目的と手法

収集した音声データに開発した評価技術を適用することで，評価技術の効果や課題，改善点を調査する．

事前準備として，4.1 節で述べたデータ収集と前処理を行った．音声データは，アナウンス初学者男性 3 名，熟練者女性 6 名，アナウンサー女性 1 名の合計 10 名のアナウンス音声であった．今回の実験で読み上げてもらった原稿を図 3 に示す．これは，筆頭著者が放送コンテスト用に自作した約 1 分半のアナウンス原稿 (390 字) である．

被験者の属性を表 3 に示す．被験者 10 人を，経験年数や大会実績を考慮し熟練度順に#1〜#10 まで設定した．被験者の中で，#2，#7，#8 の被験者は大学内でダイナミックマイクを用いて録音を行っており，その他の被験者は自宅でスマートホンのマイクで録音した．ダイナミックマイクでの録音風景を図 4 に示す．

[*2] 著者は高校・大学で放送部に所属しており，NHK 放送コンテストなどの大会出場経験あり

ブラック・ジャックの新作が公開されます。はこだて未来大学の村井源教授が参加しているプロジェクト「TEZUKA2023」は、AI を活用してブラック・ジャックの新作漫画を制作しています。3 年前に、「TEZUKA2020」で AI を利用して制作を行った手塚作品の新作「ぱいどん」では、全体の 1,2 割しか AI を活用できなかったと言います。しかし、今回のプロジェクトは近年の GPT などの発達により、半分以上の作業を AI が担っています。AI を利用し「ブラック・ジャック」における話の展開のクセを分析し、「ブラック・ジャック」らしい物語の展開の提案や、キャラクターの原画の作成、コマ割りなどを行います。AI を利用した物語制作の意義について村井教授は、「著作権侵害などが問題になっている今こそ、AI によってクリエイターが損害を被るのではなく、人間の想像力の底上げにつながってほしい」と話します。AI の力を借りて帰ってくる。「ブラック・ジャック」の新作は 2023 年公開予定です。

図 3 評価実験で利用した原稿データ

表 3 被験者の属性・経験年数・放送経験

	属性	経験年数	放送経験について
#1	初学者	ほぼなし	ほぼなし
#2	初学者	1 年未満	大学放送サークル所属
#3	初学者	1 年未満	朗読経験あり
#4	経験者	約 3 年	高校放送部所属
#5	経験者	約 3 年	全道大会入賞経験あり
#6	経験者	約 3 年	朗読部門全国大会出場経験あり
#7	経験者	約 4 年	高校・大学放送部所属
#8	経験者	約 4 年	大学放送コンテスト決勝進出経験あり
#9	経験者	約 4 年	高校放送部・アナウンススクール在籍経験あり
#10	アナウンサー	7 年以上	高校放送部所属・現役アナウンサー

図 4 大学での録音風景

5.2 結果と考察

各評価技術をアナウンス音声に適用した結果と考察について示す．

5.2.1 発声

音声データの音圧レベル (dB) の最大値・最小値・平均値・標準偏差を求めた．音圧レベルを求めるにあたって，無音部分を削除した．結果を表 4 に，属性ごとの平均と標準偏差を比較したものを表 5 に示す．表 5 の通り，初学者全体と熟練者全体を比べると熟練者の方が標準偏差が小さい．しかし，被験者ごとに見ていくと，初学者#1，#2 に関しては熟練者全体の標準偏差との差は大きくなかったため，標準偏差と熟練度が必ず対応しているとは言えない．

今回の被験者は初学者についてもアナウンス経験がわずかにあったことや，録音環境の統一を行えなかったことなどが原因として考えられるため，今後改善したい．

表 4 音圧 (dB) の最大値・最小値・平均値・標準偏差

	最大値	最小値	平均値	標準偏差
#1	-13.13	-57.03	-28.26	8.18
#2	-10.02	-65.41	-24.8	7.03
#3	-10.59	-83.97	-27.39	15.51
#4	-11.2	-47.36	-26.39	7.45
#5	-17.89	-50.08	-26.31	4.73
#6	-6.94	-57.71	-18.31	7.08
#7	-13.44	-65.16	-25.33	6.77
#8	-12.59	-43.29	-22.11	5.06
#9	-15.61	-61.67	-31.53	10.1
#10	-8.76	-53.2	-18.2	7.92

表 5 属性ごとの標準偏差

属性	標準偏差
初学者	10.24
熟練者	7.01

5.2.2 発音

文字起こしでの誤認識数を表 6 に，初学者と熟練者の平均誤認識数を表 7 に示す．

表 6 各被験者の文字起こしの誤認識数

	単語の差異数	具体的な単語
#1	3	手塚，に，ぱいどん，
#2	1	ぱいどん
#3	1	12 割
#4	1	プロジェクト
#5	4	2020，では，を，ぱいどん，
#6	0	
#7	2	2020，ぱいどん，
#8	1	プロジェクト，
#9	1	ぱいどん
#10	0	

表 7 属性ごとの文字起こしの平均誤認識数

属性	誤認識数
初学者	1.66
熟練者	1.42

初学者の方が平均誤認識数は多かったが，全体的に文字起こしの精度が高く誤認識の数に大きな違いはみられなかった．今後は音声認識だけではなく，音響分析による発音の分析も進めていきたい．

細かい点を見ていくと，誤認識が全くなかったのは#6 と#10 の経験者とアナウンサーであった．また，経験者や初学者に関係なく原稿に出てくる「ぱいどん」という単語が誤認識される場合が多く，文字起こしの機能では，固有名詞を出力させるのが難しい可能性が高いと考えられる．さらに，本来ないはずの助詞が認識されている場合もあった．人が聞いている場合は，前後の文章から助詞の有無や，正しい助詞を聞き取ることが出来るが，音声認識による文字起こしでは，若干音が伸びていたりすると助詞として認識されることが分かった．今後，こうした特徴を考慮したうえで発音判定機能の改良を進めていく．

5.2.3 アクセント

今回用意した原稿のアクセントを NHK 日本語アクセント新辞典から引用した．アクセント記号の説明を表 8 に示す．原稿のアクセント表記の一覧を表 9 に示す．無声音については，今回分析対象としていないため表記から省略する．

表 8 アクセント記号

音の下がり目	＼
平坦（音）	＿
鼻濁音	゜

表 9 原稿のアクセント表記一覧（無声音は除外）

単語	アクセント表記	単語	アクセント表記
ブラック	ブラ＼ック	分析	ブンセキ＿
新作	シンサク＿	物語	モノカ゜＼タリ
公開	コーカイ＿	提案	テーアン＿
○○大学	○○ダ＼イカ゜ク	キャラクター	キャラ＼クター
○○教授	○○キョ＼ージュ	原画	ゲンカ゜＿
参加	サンカ＿	作成	サクセー＿
プロジェクト	プロ＼ジェクト	など	ナ＼ド
活用	カツヨー＿	行い	オコナイ＿
制作	セーサク＿	意義	イ＼ギ
三年	サンネン＿	著作権	チョサ＼クケン
○○前	○○マ＼エ	侵害	シンカ゜イ＿
利用	リヨー＿	今	イ＼マ
全体	ゼンタイ＿	損害	ソンカ゜イ＿
出来ない	デキ＼ナイ	被る	コーム＼ル
しかし	シカ＼シ	人間	ニンゲ゜ン＿
今回	コ＼ンカイ	創造力	ソーゾ＼－リョク
近年	キ＼ンネン	底上げ	ソコアゲ゜＿
発達	ハッタツ＿	繋がって	ツナカ゜ッテ＿
半分	ハンブ＼ン	話し	ハナ＼シ
以上	イ＼ジョー	力	チカラ＼
作業	サ＼ギョー	借りて	カリテ＿
担う	ニナ＼ウ	帰って	カ＼エッテ
話	ハナシ＼	予定	ヨテー＿
展開	テンカイ＿		
くせ	クセ＼		

被験者の録音音声を分割し，上記のアクセントと照合する．例として，被験者#1 と被験者#10 のピッチ (f0) とアクセントを図 5，図 6 に示す．画像の下にあるのが，音声に対応した文字で，青い矢印が音の下がり目を示している．それに対し，図 5 の赤い丸の部分に注目すると，音が上がっているのが分かる．この場合，被験者#1 のブラックのアクセントは間違っていることになる．それに対し，図 6 の被験者#10 は f0 に注目すると，「ブラック」の “ラ” の部分で音が上がり，“ッ” の部分で音がやや下がっている．

このような手法を用いて，著者が目視で，被験者#1，#5，#10 に対して録音データ全体のアクセントの正誤を判定した．初学者の中でも特にアナウンス経験がない#1，経験者の中でもアクセントやイントネーションを比較的正しく表

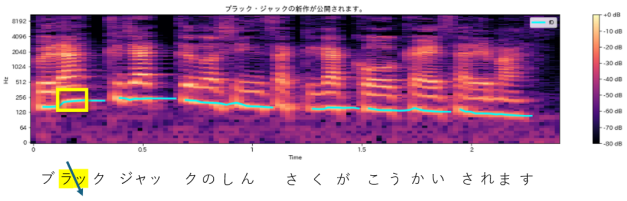

図 5 被験者#1 のピッチ

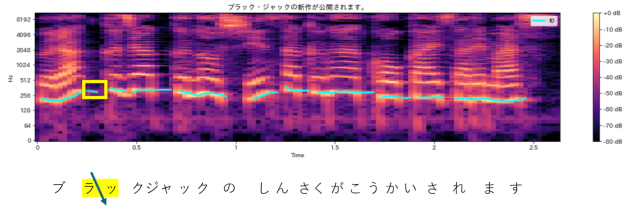

図 6 被験者#10 のピッチ

現できている#5，本職のアナウンサーの#10 の 3 人を代表として選出した．結果を表 10 に示す．

表 10 被験者#1，#5，#10 のアクセントの正解率

被験者	#1	#5	#10
正解率	66.7%	64.6%	72.9%

アクセントの正解率についてはあまり熟練度の関係性が見られなかった．これは，熟練者でも全てのアクセントを熟知しているわけではなく，原稿のアクセントを確認してから読む場合がほとんどであるため，今回の実験のように同じ原稿をアクセントを調べずに読むとアクセントの正解率にあまり差が見られないと考えられる．

5.2.4 イントネーション

アクセントと同様に，被験者#1，#5，#10 のイントネーションについて特徴を比較する．被験者#1，#5，#10 の原稿全体のピッチを可視化したものを図 7 に，20〜30 秒部分を拡大したものを図 8 に示す．

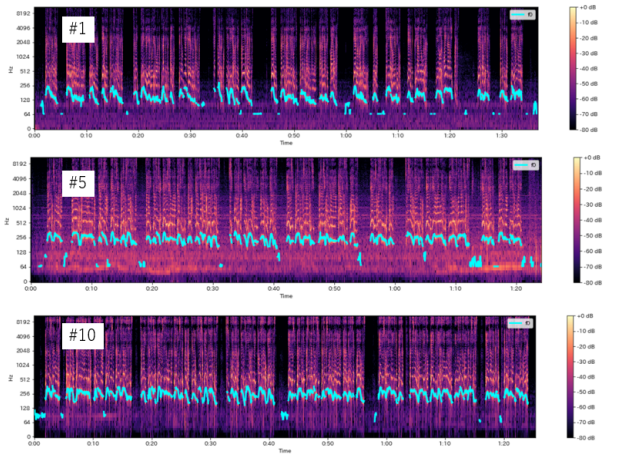

図 7 被験者#1，#5，#10 の音声全体のピッチ

初学者#1 については，文末では音が下がりきっており，音を一定に下げる「下し」という技術は使えていることが

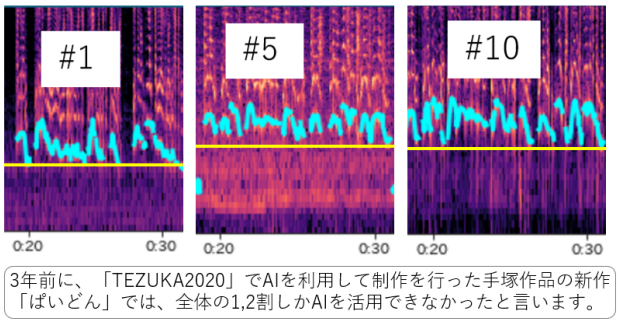

図 8 被験者#1，#5，#10 のピッチの拡大図 (20〜30 秒)

分かる．一方，音の下がり方がどの文も一定であった．また，経験者の#5 は文末で音を下げきれていると同時に，被験者#1 に比べ音の下げ方にばらつきがあり，表現の多様性が感じられた．一方，アナウンサーの被験者#10 は音を下げきれていないところが多かった．放送コンテストはしっかり音を下げるのに対し，現役のアナウンサーは聴き手に明るい印象を与えるために語尾を上げる傾向がある可能性が考えられる．

今後は，大会音声データなどから熟練者の音の下げ方の傾向を分析し，イントネーションの指標を再考していきたい．

5.2.5 テンポ

文字起こしした原稿の文字数と録音した音声の長さから，1 分あたりに読んでいる文字数を計算する．今回の実験では被験者全員に同じ原稿を読んでもらったため，文字数を原稿通りの 390 文字に統一した．音声ファイルの長さと文字数から計算した結果を表 11 に示す．

表 11 被験者ごとの原稿 (390 字) を読む速度

	1 分あたりの字数	属性ごとの平均 (標準偏差)
#1	241.1	
#2	250.2	
#3	276.7	256(18.4)
#4	280.1	
#5	278.1	
#6	271.3	
#7	276.3	
#8	283.4	
#9	245.2	
#10	273.9	272.6(12.7)

熟練者の方が分間 300 字程度の目安に近い値となり初学者の方が読むスピードが遅いと考えられる．しかし，分間 300 字に達している被験者はいなかったことから，大会音声データからも熟練者の話速度を収集し目安の見直しを行っていきたい．

5.2.6 ポーズ

実装した inaSpeechSegmenter ライブラリによる無音時間の検知では，句点での無音時間は確実に検知出来ていた

ものの，被験者 10 人中 6 人の読点を認識出来なかった．読点での無音時間も検知できるよう今後検討を進めていく．

上記を踏まえ，各被験者の句点のみの無音時間の秒数を分析した．なお，最後の句点での無音時間は録音終了のタイミングにより差異が出るため，分析には含めない．各被験者の句点での無音時間の平均秒数と標準偏差を表 12 に示す．

表 12 句点でのポーズ

	平均 (標準偏差)	属性ごとの平均 (標準偏差)
#1	2.52(0.81)	
#2	1.96(0.92)	
#3	2.22(0.76)	2.23(0.83)
#4	1.44(0.07)	
#5	1.35(0.4)	
#6	2.05(0.59)	
#7	1.94(0.89)	
#8	1.78(0.57)	
#9	1.7(0.52)	
#10	1.34(0.71)	1.65(0.53)

表 12 から，初学者の方が句点での間が長く，長さのばらつきも大きいということが分かった．また，熟練者の句点での間の平均が 1.65 秒と目安に設定していた 2 秒より小さいという結果になったため，句点の間の長さの目安の見直しも合わせて検討していきたい．

6. まとめと展望

本研究では，高校や大学で行われる競技アナウンスの採点基準を参考にして，アナウンス技術の評価指標を考察した．次に，アナウンス初心者／経験者の音声データを収集し，評価指標に基づいて，アナウンス技術を分析するシステムの実装を進めた．さらに性能評価では，10 人の被験者からアナウンスの音声データを収集し，設定した評価指標や評価技術を検討した．

一方，現状のシステムには技術的な課題が多く残っているので，その改善を進める．また，研究の最終目標はアナウンスの学習支援システムであるため，フィードバック手法の開発を行う．加えて，大会上位者のアナウンス音声の収集を行うことで評価指標をより正確なものにする．

参考文献

[1] 第 70 回 NHK 杯全国高校放送コンテスト「要項号」, https://www.nhkfdn.or.jp/kyoiku/ncon/ncon_h/pdf/70/70_youkou.pdf, (参照 2023-06-25）.
[2] 第 40 回 NHK 全国大学放送コンテスト要項, https://drive.google.com/file/d/18lIMIX3z1oDr_XM5PwYRZQct0MLZxN8l/view?usp=drive_link,（参照 2023-06-25）.
[3] 中野 倫靖，後藤 真孝，歌声インタフェース：歌声を対象とした信号処理とそれに基づくインタフェース構築，日本音響学会 2013 年秋季研究発表会 講演論文集，3-7-3, pp. 1465-1468(2013).
[4] 羽賀翼，香山瑞恵，池田京子，橋本昌巳，伊東一典，習熟度に関係する音響特徴量に基づく 歌唱学習支援システムの評価，人工知能学会第二種研究会資料，身体知研究会，SIG-SKL-20-01(2015).
[5] 栗原一貴，後藤真孝，緒方淳，松坂要佐，五十嵐健夫，プレゼン先生：音声情報処理と画像情報処理を用いたプレゼンテーションのトレーニングシステム，WISS2006 論文集，NO. 43，pp59-64(2006).
[6] 趙新博，由井薗隆也，宗森純，初心者向けプレゼンテーション練習支援システム PRESENCE の開発と評価，電気学会論文誌 C，Vol.138，NO.10，pp1269-1277(2018).
[7] IAM, K.K.，声優滑舌アプリ，https://apps.apple.com/jp/app/%E5%A3%B0%E5%84%AA%E6%BB%91%E8%88%8C%E3%82%A2%E3%83%97%E3%83%AA/id642099721,（ 参照 2023-10-30）.
[8] Space Factory Inc., SAY-U, https://apps.apple.com/jp/app/say-u/id1533211139,（参照 2023-10-30）.
[9] 王 伸子，日本語教育の教材としての音声素材の音響的分析 ―ナレーション、アナウンス、声優ボイスオーバーの分析― ，専修大学外国語教育論集，50 57-73,(2022).
[10] ＮＨＫ放送文化研究所，日本語発音アクセント新辞典，NHK 出版，2016.
[11] 矢野香，【第 2 回】「ＮＨＫ式 7 つのルール」をマネれば、あなたの話し方が一変する！,DIAMOND online, https://diamond.jp/articles/-/56483?page=3, (参照 2024-01-23).