

歌による嚥下・構音機能の定量的評価手法の実現に向けた 歌唱データの音響・画像分析

平井 雄太^{1,a)} 耿 世嫻¹ 下島 銀士¹ 小野寺 宏¹ 矢谷 浩司¹

概要：近年、高齢者の軽微な口腔機能の低下が将来の健康リスクを増加させることが示され、口腔機能の定期的な評価の重要性はますます高まっている。しかしながら既存の口腔機能の評価手法は、特別な測定機器の必要性や測定自体の単調さ等の理由から、ユーザーが自発的に繰り返し取り組み難いという課題がある。そこでカラオケに含まれるゲーム性が測定に対する心理的負担を軽減することに着目し、モバイル端末による歌唱を通じた嚥下・構音機能の評価手法を提案する。提案手法の実現に向けて、我々は非高齢者110名（65歳未満）と高齢者76名（65歳以上）の実験参加者から歌唱と構音・嚥下機能のデータを収集した。本稿では、得られた歌唱データに対して音響・画像分析を行い、構音・嚥下機能との関連を示した後、提案手法の実現に向けた今後の課題について検討する。

1. はじめに

フレイルとは、主に高齢者が健康な状態と要介護状態の中間の状態にあることを示す概念で、適切な介入によって再び健康な状態に戻ることのできる可逆性がその大きな特徴である。近年、口腔機能の軽微な低下に伴う食習慣の悪化がその後の要介護状態や筋量低下などのリスクを増加させ、死亡率に至っては健康な場合と比較して約2倍にも達することが示された [14]。そこで口腔機能の維持の重要性を啓発するため、「オーラルフレイル」と呼ばれる新概念が提唱され、注目を集めている。フレイルと同様に、オーラルフレイルも適切な介入を行えば再び健康な状態に戻ることができるため、口腔機能の低下を早期に発見し、リハビリテーションを行っていくことは極めて重要である。

口腔機能の評価する既存手法として、/pa/, /ta/, /ka/, /ra/をそれぞれ5秒間に発音できた回数で口腔機能の評価するオーラルディアドコネシス [18] や、何回連続で唾液を嚥下できたかによって口腔機能の評価する反復唾液嚥下テスト [17] が挙げられる。こうした既存の口腔機能の評価手法は手軽である一方、医療従事者による観察を必要とする点や、手法自体が単調であるため、自発的・継続的に取り組むには心理的な負担が大きい点が課題として挙げられる。また、舌圧の測定等の評価手法は体内に測定器を挿入するなど侵襲的であるため、被測定者に身体的な負担が生じることがある。

そこで我々は上記の課題を解決するため、モバイル端末を用いた、歌唱による口腔機能の定量的評価手法を提案する。提案手法のコンセプトを図1に示す。これはいわば、カラオケの歌唱力採点機能に口腔機能の採点機能を加えたシステムである。利用者はスマートフォンなどのモバイル端末から流れる伴奏に合わせて歌唱を行い、その様子をモバイル端末に搭載されたマイクとカメラによって記録する。曲が終了すると、我々の提案するアルゴリズムに従って、口腔機能の点数が表示される。提案手法は、医療従事者の助けを必要とせず、そして非侵襲的に口腔機能の評価することができるだけでなく、歌唱とその採点というゲーム性を含めることによって、利用者の自発的・継続的な活用が期待できる。

提案手法の実現に向けて、本稿では我々が収集した非高齢者及び高齢者の歌唱時の音響・画像分析を行い、嚥下・構音機能との関連性を調査する。我々が収集したデータには、歌唱時の音声・動画に加えて、アンケート調査による嚥下機能の評価手法である EAT-10 ^{*1} と定量的な構音機能の評価手法であるオーラルディアドコネシス試験時の音声・動画が含まれている。収集した歌唱データから音響・画像特徴量を計算したのち、変数増減法によるロジスティック回帰分析によって嚥下・構音障害の分類に有用な特徴量の選択を行った。そして、選択された特徴量を用いたロジスティック回帰による嚥下・構音障害の分類性能について評価を行なった。

¹ 東京大学
^{a)} hirai@iis-lab.org

^{*1} <https://www.nestlenutrition-institute.org/resources/nutrition-tools/details/swallowing-assessment-tool>

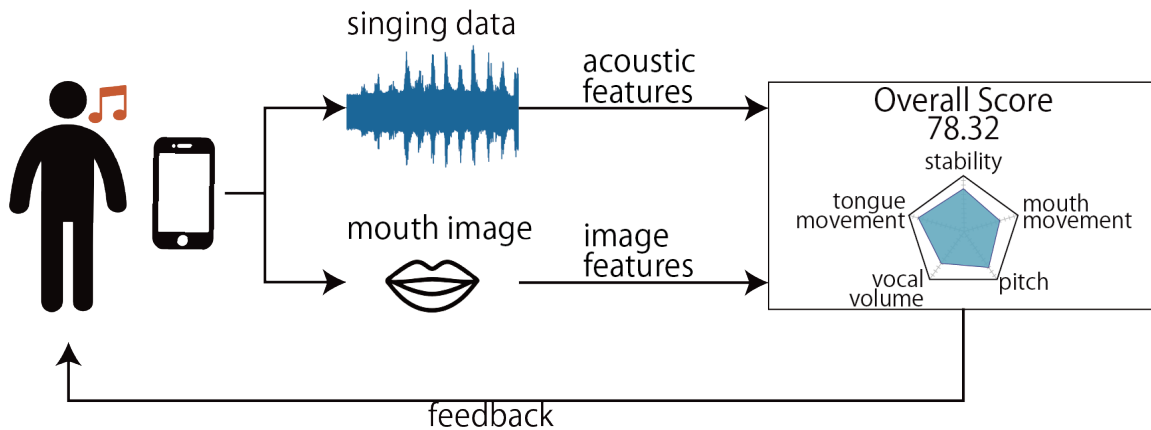


図 1 提案するシステムのコンセプト図. ユーザーはスマートフォンから流れる音源に従って歌唱を行い, その様子の録画を行う. 歌唱が終了すると, 録画の結果から音響・画像特徴量が抽出され, 採点結果を表示する.

2. 関連研究

2.1 オーラルフレイル

フレイルとは特に高齢期にみられる身体機能の低下のことであり, 将来の転倒, 障害, 要介護, 死亡などのリスクと大きく関連する一方で, 健康状態と要介護状態の間の中間的な状態にあり, 適切な介入により心身の機能を回復させることができるという特徴を持つ [5].

近年, 咀嚼能力の低下や天然歯の減少などの軽微な口腔状態の低下が, 将来のフレイルや要介護状態, 死亡率などのリスクの増加と関連していることが示された [14]. そこで, 歯科口腔系の軽微な機能低下を医療従事者や一般国民が見逃さないようにするため, 「オーラルフレイル」という概念が提案された [21]. オーラルフレイルもフレイルと同様に, 適切な措置を取ることで健康な状態に戻ることができるため, 早期にこれを発見して適切な措置を取り, 将来の健康リスクを減らすことが社会的課題となる.

2.2 口腔機能の評価手法

口腔機能の定量的な評価手法はこれまで数多く提案されており, 評価項目として口腔衛生状態, 舌圧, 吻合力, 舌口唇運動機能, 咀嚼・嚥下機能等が一般的である [20]. それらの中でも本研究との関連が深い舌口唇運動機能と嚥下機能の評価手法について本節で説明する.

舌口唇運動機能の評価手法として広く用いられているオーラルディアドコキネシスは, /pa/, /ta/, /ka/, /ra/ を一定時間内に繰り返し発音できた回数を指標とする [18]. 発音回数の基準値は各医学会, 医師会などが独自に定めていることが多い. 発音回数は医師が手動で計測するほか, 計算機を用いて自動で回数を測定する方法も提案されている [13, 16].

嚥下機能のスクリーニングツールとして用いられている

のが EAT-10 である [2]. EAT-10 では, 例えば「飲み込みの問題が原因で, 体重が減少した」などの計 10 個の質問に対して 0 4 点で主観的評価を記入する質問用紙であり, 合計点数が 3 点以上の場合には結果を専門医に相談することが推奨される. 日本を含む世界各国で嚥下機能を評価する手法として有効であることが示されている [3, 12].

2.3 口腔機能の訓練を目指したアプリケーション

口腔機能低下予防に向けた取り組みとして「口腔体操」が各自治体で採用され, 日本医師会によるオーラルフレイル対策にも取り上げられている [19].

口腔機能向上トレーニングは様々に提案されている一方で, 高齢者自身が自発的・継続的に取り組むには課題が多い. そこでゲームの要素を組み入れて自発的・継続的なトレーニングの促進を目指したのが Squach である [1]. Squach は口及び舌の動きを使って操作するビデオゲームで, 介護施設入居者を対象とした実験では, オーラルディアドコキネシスなどいくつかの口腔機能の評価において改善が報告されている.

2.4 構音障害の音響分析

構音機能の評価には, 患者の発声の分析が用いられることが多い. 例えば, セラピストが患者の発声の評価する際の基準として GRBAS 尺度などが知られ, 一定の信頼性が認められている [4]. 一方でより客観的でコストのかからない, 計算機による評価指標の確立に向けて, これまで様々な音響特徴量が提案されてきた.

構音障害を評価する古典的な音響特徴量として, jitter, shimmer, harmonic-to-noise ratio (HNR), メル周波数ケプストラム係数 (Mel Frequency Cepstral Coefficient, MFCC) などの音響特徴量を用いる手法が挙げられる. これらは, 数秒間持続した 1 語の発音に対してしばしば計算

される特徴量である。jitter, shimmer, HNR は、発声中の音声が発音者では周期的であるのに対し、構音障害者では非周期的であることに着目した特徴量である。jitter は発声中の基本周波数の偏差を、shimmer は振幅の偏差を定量化し、HNR は周期成分と非周期成分の比率を表す。実際に jitter, shimmer, HNR によって健常者と構音障害者を分類可能であることが示されている [11]。一方で、jitter, shimmer, HNR の算出に必要な基本周波数は、構音障害者を対象とした場合に正しく計算できない可能性があるため、基本周波数を必要としない MFCC などの特徴量を用いて、構音障害を検出する手法も提案されている [7]。

こうした手法に加え、非周期成分に着目した Recurrent Period Density Entropy (RPDE) [11], detrended fluctuation analysis (DFA) [8] などこれまで様々な特徴量が提案されている。

3. データ収集と分析手法

3.1 データ収集

歌による嚙下・構音機能の定量的評価手法の実現に向けて、我々はまず、国内のクラウドソーシングサービスの CrowdWorks^{*2}と Craudia^{*3}で実験参加者を募集し、109 件のデータを得た。クラウドソーシングで募集した参加者の年齢を調べた結果、図 2 (a) に示すようにほとんどが 60 代未満であった。我々がめざすアプリケーションの主な対象者である高齢者のデータを収集するため、次に我々は東京都文京区で 60 歳から 90 歳の計 76 名の実験参加者を募集した。なお、その際のデータ収集はオンラインでは実施せず、実験は全て同一施設で行い、モバイル端末の操作は著者らが行った。

実験は以下の手順で実施した。まず、google forms を用いたアンケート調査を行い、性別・年齢の情報を得た。次に、嚙下機能の評価するために EAT-10 を用いたアンケート調査を行い、例えば「飲み込みの問題が原因で、体重が減少した」などの合計 10 個の質問に対して、「0=問題なし」から「4=ひどく問題」の 5 段階で主観的に評価してもらった。

その後、我々が作成したウェブサイト上で以下の実験を実施した。まず、/pa/, /ta/, /ka/, /ra/ の語をそれぞれ 5 秒間できるだけ多く繰り返し発音するオーラルディアドコキネシス試験を行い、その様子の録音と録画を行った。なお、プライバシーを考慮し、実験中は図 3 (b) のように、鼻と口が映り、鼻より上の部分が写らないようにカメラの位置を調整するようお願いした。

次に、これらの嚙下・構音機能と歌唱時のデータを結びつけるため、歌唱とその様子の録音・録画を行った。撮影範囲はオーラルディアドコキネシスの時同様に、鼻より

下の範囲とした。歌唱する歌には、認知度の高い曲として「ふるさと」(高野辰之作詞・岡野貞一作曲)を選択した。実験を開始すると、スマートフォンから伴奏が流れ始め、図 3 (a) に示すように、その時点で歌うべき箇所が赤色で強調して表示される。

システムの不具合によりファイルサイズが 0 Byte であるものを含む参加者のデータ、いくつかの実験を正しく行わなかった参加者のデータ、正しく録音・録画ができなかったと申告のあった参加者のデータを除外し、クラウドソーシングで 100 件、シビックセンターで 75 件のデータを得た。「ふるさと」の歌唱の際、実際の歌唱前に数秒間ラグが入る場合があったため、このラグをあらかじめ消去し、どのデータの歌唱時間も 45 秒間になるように調整した。

3.2 嚙下・構音障害の基準

今回、機械学習による嚙下・構音障害の分類を行うにあたって、嚙下・構音機能障害の基準を定める必要がある。本稿では、医師への相談が推奨される、EAT-10 スコアの合計が 3 点以上を嚙下障害と定義した。また、構音障害は /pa/, /ta/, /ka/, /ra/ の 5 秒間の発音回数のいずれかが 15 回以下であることとした。

3.3 オーラルディアドコキネシスの音響分析

まず、オーラルディアドコキネシスに対する音響分析の結果が先行研究と一致するかどうかを確かめるため、各発音を音節ごとに分割して、特徴量を計算することで健常者と構音障害者を計算機によって区別する先行研究を再実装し、評価を行った [13]。図 4 に示すように、この手法では音声信号を 1 音節ごとに分割したのち、initial burst (空気流に起因する発音の開始地点), vowel onset (声帯の振動に起因する周期的な音声信号の開始地点), occlusion (緩やかな音声信号の減少) を求めて、特徴量を計算する。なお、再実装にあたっては、著者らが提案したベイズ変化点検知手法の代わりに、より実装の容易な 2 元変化点検知を用いた [6]。特徴量には、/pa/ の連続発音時の vvq (vowel variability quotient, 1 音節中における振幅の分散), vot (vowel onset time, initial burst から vowel onset までの時間), cst (consonant spectral trend, スペクトラムの線形回帰係数) の 3 つを選択し、これを説明変数とした。また、目的変数は嚙下障害および構音障害とした。そして、SVM (support vector machine) による 5 分割交差検証によって音響特徴量の有用性を評価した。

3.4 歌唱の音響分析

我々はまず、前節と同様の手法により、initial burst, vowel onset, occlusion の検出を行い、特徴量の計算を試みた。しかしながら、/pa/ などの特定の語を連続で発音するオーラルディアドコキネシスと違い、歌唱中は音節ごと

*2 <https://crowdworks.jp/>

*3 <https://www.craudia.com/>

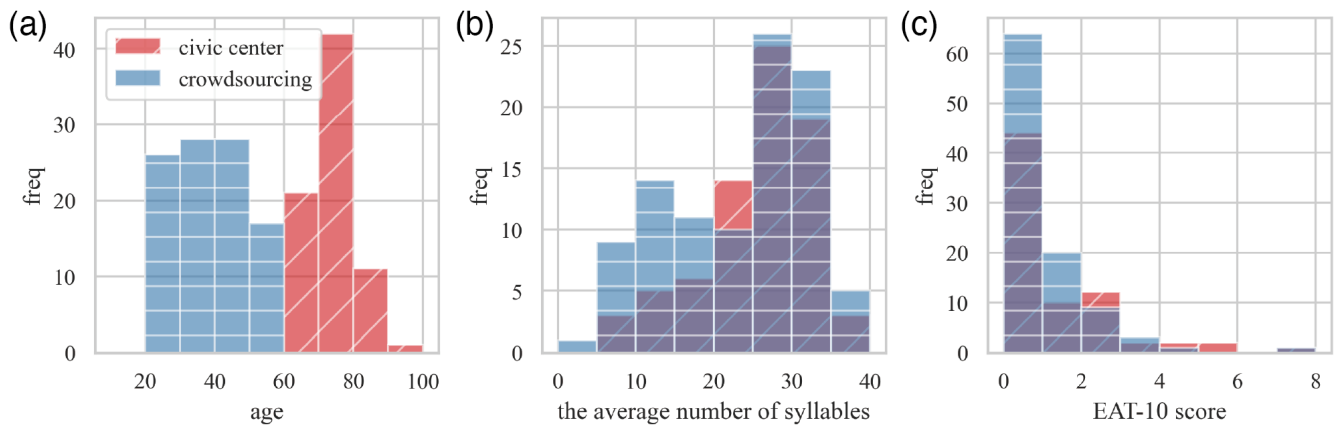


図 2 参加者の分布を示したヒストグラム. 青がクラウドソーシングでの、赤が東京都文京区の高齢者の参加者を表す. 左から順に, (a) 年齢, (b) /pa/, /ta/, /ka/, /ra/ の 5 秒間の発音回数の平均, (c) EAT-10 スコアを表す.

(a) ふるさと

$\text{♩} = 80$

(b)

図 3 「ふるさと」の歌唱の実験のため使用したウェブサイトのページ (抜粋). (a) 歌唱中は「ふるさと」の伴奏が流れ, 歌うべき箇所が楽譜上で赤く表示される. また, スマートフォンのフロントカメラにより, (b) のように映像が表示される. 被験者には鼻を含めた鼻より下の範囲が映るようにカメラの位置を固定するようお願いをした. 実際に実験に使用したウェブサイトでは説明文が書かれているほか, 一部レイアウトが図とは異なる.

表 1 歌唱データの音響分析に用いた特徴量

名称	説明
stdF0	基本周波数の標準偏差
hnr	音声の周期性の比率
jitter	基本周波数の変動する割合
shimmer	振幅の変動する割合
cpp	ケプストラムのピーク値の卓立度
dfa	DFA によって計算された Hurst 係数 [8]
dfa_norm	dfa をシグモイド関数に代入した値

の境界が明確ではないため, この手法はうまく機能しないことが分かった.

そこで, 歌唱データが伴奏に合わせて歌われたものであ

るため, 開始から何秒の時点でどの部分を歌唱したのかをおおよそ知ることができることに着目した. まず, 図 3 (a) の楽譜の歌詞がある音符ごとに録画データを分割し, これらを音節とした. その際, 実際の歌唱は, 本来歌うべきパートからずれることがあることを考慮して, 境界の前後に 100 ms のマージンを加え, 本来の時間から 200 ms 短くなる用にした. 次に, 分割された各音声に対して, 表 1 に示す特徴量を計算し, 特徴量ごとに標準化を行った.

そして, 得られた特徴量に対して以下の手順で変数増減法によるロジスティック回帰分析を行い, 有用な特徴量の選択を行った. まず, 選択された特徴量がゼロの状態からスタートし, 特徴量を加えたときに赤池情報量規準 (AIC)

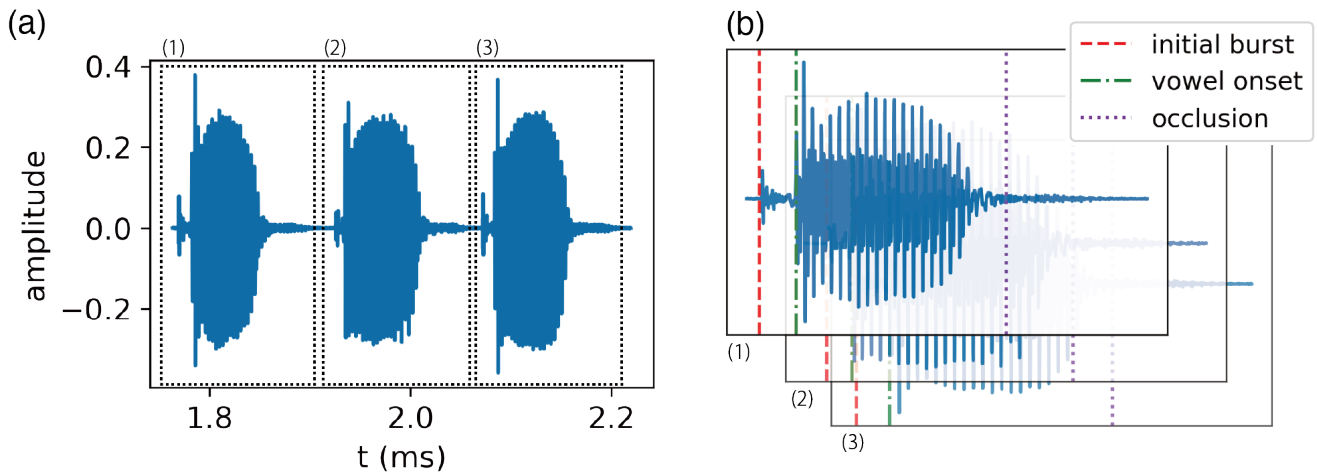


図 4 オーラルディアドコキネシス時の連続した発音の分析手法 [7]. (a) 同じ語を繰り返し発音した時の音声を音節ごとに分割する. (b)(a)において分割された各音節ごとに, initial burst, vowel onset, occlusion の箇所を計算する.

が減少する特徴量を有用な特徴量として選択する. ただし, 多重線形性を回避するため, 分散拡大係数が5より大きくなる特徴量は除いた. これを, AIC がそれ以上小さくなくなるか, 選択された特徴量の数が10個に達するまで繰り返した. その後, 得られた特徴量による嚙下・構音障害の分類性能をロジスティック回帰によって確かめた.

3.5 歌唱の画像分析

構音・嚙下機能の評価に向けた画像分析は未だ未開拓の領域であり, どのような特徴量がそれらの予測に適しているのかも不明である. そこで, 我々は撮影画像から口の高さ・幅と口の大きさに対して, 時系列データ一般に適用される特徴量を歌唱中の動画データに対して適用し, 有用な特徴量を選ぶことにした.

まず, 口の高さ・幅と口の大きさを歌唱中の録画データから推定する手順を図5に示す. まず, 図5(a)のように, 収集した歌唱中の動画データをフレームごとに分割して画像データとし, haar cascade を用いて口の範囲を抽出し, その範囲を中心に縦と横の大きさを3倍に拡大した後, 縦128px, 横256pxになるようにリサイズした [15]. ただし, haar cascade では正しく口の範囲を推定できていないことが多かった. 本稿の目的は, 完全なシステムの構築ではなく, 歌唱と嚙下・構音機能の関係の調査であるため, うまく口の範囲が切り出せていない画像は手動でラベリングし, その画像の口の範囲は, 正しい前後の切り出し範囲と同じとした.

次に, haar cascade によって切り出された口の画像のどのピクセルが口に対応しているかを調べるため, 近年提案された MaskGan を用い, 顔画像の各ピクセルを顔の部位ごとに分類する face segmentation を行った [9]. MaskGan では, 顔画像と手動でラベリングされた正解データをデー

ラーニングにより学習することで face segmentation を行う. MaskGan に用いられたデータセットである CelebA-HQ には, 正面を向いた人の顔画像が含まれており, 下半分のみを切り出すことで, 鼻より下の範囲のみが概ね映るようにすることができる. そこで, 鼻より下の部分のみを含む我々のデータに対して MaskGan を適用するため, CelebA-HQ の各画像の下半分のみを用いて再学習を行った. MaskGan による顔の部位の分類結果を図5(b)に示す. 本稿では, MaskGan による face-segmentation の結果出力される顔の部位のうち, 上唇, 下唇, 口の中の部分を口とみなした. 図5(c)のように, 口に分類された部分のみを白く表示すると, 本来口ではないピクセルの一部が誤検出されていることがわかる. 誤検出を排除するため, 隣り合う各ピクセルを連結した時に, 総ピクセル数が最大になる部分のみを口であるとした. 最終的に口に分類された部分を白く表示した画像を図5(d)に示す. この時の高さ・幅・口の面積(口の総ピクセル数)を動画の各フレームに対して計算した.

その後, 前節と同様に歌詞ごとにデータを分割したのち, 時系列データ分析のための Python のライブラリである tsfresh^{*4}を用い, 高さ・幅・口の面積に対する特徴量の列挙を行った. そして, 前節の音響特徴量の選択手法と同様の変数増減法により, 有用な特徴量を選択し, ロジスティック回帰による分類性能を確かめた.

4. 分析結果と考察

4.1 オーラルディアドコキネシスの音響分析

SVM による5分割交差検証による構音・嚙下機能の予測結果を表2に示す. 構音障害の予測では, 正解率85%, 適合率93%, 再現率65%となり, 構音障害を概ね分類可能

*4 <https://github.com/blue-yonder/tsfresh>

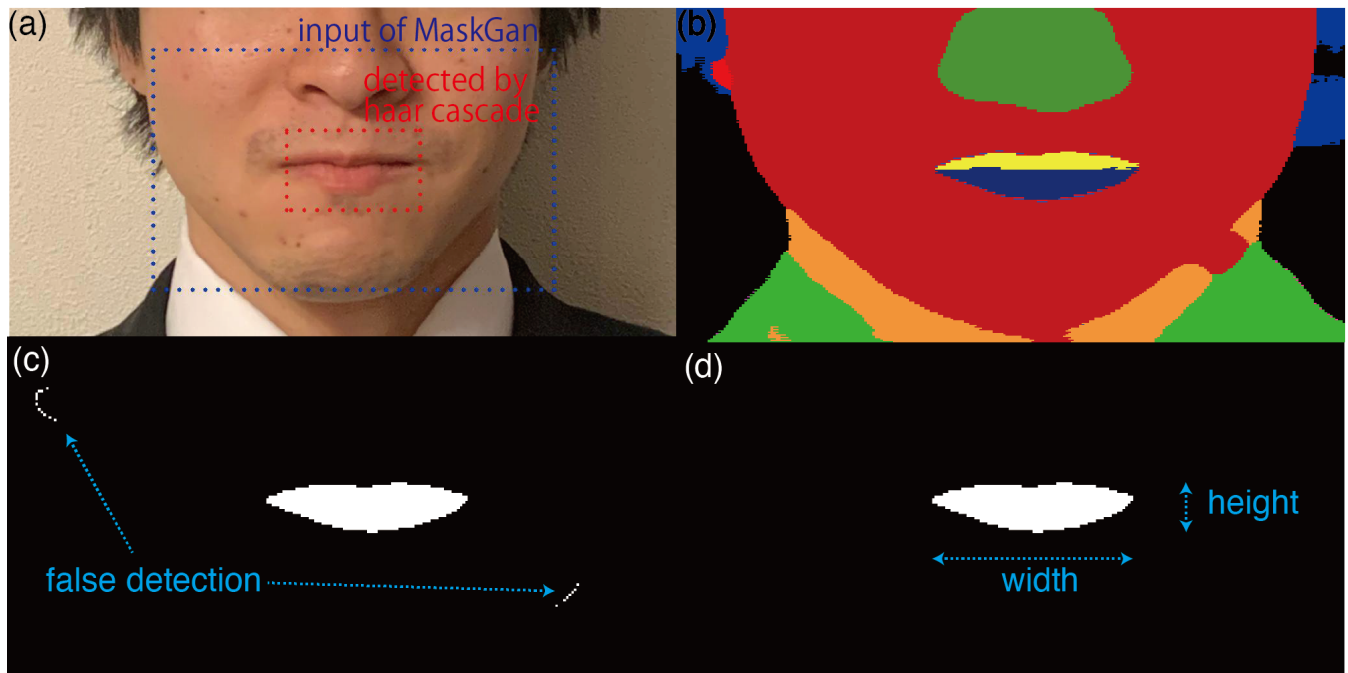


図 5 画像中の口の部分を推定手順. (a) まず, haar cascade を用いて口の範囲を推定し, 縦横比が 1:2 になるように調整する (図中赤点線) [15]. その後, 縦・横にそれぞれ 3 倍に拡大した範囲を MaskGan の入力とする (図中青点線) [9]. (b)MaskGan によって出力された face segmentation の結果. 顔の部位ごとに色分けをしている. (c)(b) の内, 上唇・下唇・口の中のみを白で表示した画像. 一部がご検出されていることが確認できる. (d) 口に分類されたピクセルのうち, 隣り合う各ピクセルを連結したときに総ピクセル数が最大になるもののみを口とし, 白で表示した画像.

表 2 /pa/の連続発音から構音・嚙下障害を予測した時の混同行列

		Predicted				Predicted	
構音障害		P	N	嚙下障害		P	N
Actual	P	34	18	Actual	P	0	15
	N	4	118		N	0	159

であることが分かった. これは先行研究と一致する [13]. 一方で, 嚙下障害の予測では全てが陰性だと判定され, 適合率, 再現率がともに 0%となり, 今回用いた特徴量は嚙下障害の予測には有用でないことが分かった.

4.2 歌唱の音響分析

変数増減法によって得られた音響特徴量を表 3, 表 4 に示す. また, これらの特徴量を用いたロジスティック回帰分析によって得られた混同行列を表 5 に示す. 嚙下障害では正解率 92%, 適合度 67%, 再現率 13%となり, 特に再現率の低さが目立つ. 一方で構音障害の分類の場合, 正解率 76%, 適合度 67%, 再現率 42%となり, 再現率を比較して, 嚙下障害より高い分類性能が得られた.

4.3 歌唱の画像分析

音響分析と同様に, 変数増減法によって得られた特徴量を表 6, 表 7 に示す. 表 6 の height_autocorrelation は口の高さの自己相関を, height_large_std は口の高さの最

表 3 変数増減法により選択された構音障害を予測する音響特徴量 括弧内は音符の番号と歌詞を表す (以下同様).

feature	coef	std err	$P > z $
const	-1.1820	0.218	0.000
hnr(16_shi)	1.1958	0.312	0.000
jitter(25_tu)	0.4506	0.207	0.030
hnr(13_gi)	-0.9857	0.339	0.004
cpp(23_bu)	-0.5643	0.246	0.022
cpp(20_ma)	0.7434	0.242	0.002
cpp(53_ta)	-0.5836	0.224	0.009
hnr(12_sa)	0.6716	0.326	0.040
jitter(39_ma)	-0.6282	0.257	0.015
shimmer(50_su)	0.4384	0.208	0.035
cpp(44_u)	-0.4147	0.233	0.075

大と最小の差と標準偏差との比率と表す. また, 表 6 の width_lempel_complexity は変化の複雑度を定量化した値であり [10], area_change_quantiles は口の面積の四分位範囲の変化の度合いを定量化した値である. これらの特徴量を用いたロジスティック回帰分析によって得られた混同行列を表 8 に示す. 嚙下障害では, 正解率 94%, 適合率 100%, 再現率 29%となっており, 音響分析の場合と同様にほとんどが嚙下障害者ではないと分類されている. 一方構音障害では, 正解率 78%, 適合率 77%, 再現率 40%となり, 嚙下障害と比較すると, 特に再現率に関して良い

表 4 変数増減法により抽出された嚥下障害を予測する音響特徴量

feature	coef	std err	$P > z $
const	-6.6844	1.406	0.000
dfa_alpha(36_ha)	-1.7405	0.643	0.007
dfa_alpha_norm(45_ri)	4.1206	1.639	0.012
cpp(38_i)	2.5413	0.702	0.000
cpp(20_ma)	-3.0874	0.904	0.001
cpp(13_gi)	1.8324	0.581	0.002
dfa_alpha(47_te)	-2.1185	0.752	0.005
dfa_alpha_norm(14_o)	2.5197	0.872	0.004
dfa_alpha(43_gu)	-2.8262	1.103	0.010
hnr(16_shi)	-1.4880	0.571	0.009
hnr(46_i)	1.0126	0.499	0.042

表 5 歌唱中の音響特徴量による構音・嚥下障害予測時の混同行列

	Predicted			Predicted			
	P	N		P	N		
Actual 構音障害	P	22	30	Actual 嚥下障害	P	2	13
	N	11	111		N	1	158

表 6 変数増減法により抽出された嚥下障害を予測する画像特徴量

feature	coef	std err	$P > z $
const	-4.1361	0.705	0.000
height_autocorrelation(14_o)	1.8895	0.480	0.000
height_large_std(50_su)	-1.4336	0.392	0.000

表 7 変数増減法により抽出された構音障害を予測する画像特徴量

feature	coef	std err	$P > z $
const	-1.0912	0.202	0.000
width_lempel_complexity(26_ri)	1.3178	0.313	0.000
area_change_quantiles(43_gu)	6.2685	2.526	0.013

表 8 歌唱中の画像特徴量による構音・嚥下障害予測時の混同行列

	Predicted			Predicted			
	P	N		P	N		
Actual 構音障害	P	20	31	Actual 嚥下障害	P	4	10
	N	6	114		N	0	157

結果が得られた。

4.4 分析結果のまとめ

オーラルディアドコキネシスに対する音響分析は構音障害をよく予測するという既存研究と一貫性のある結果が得られた。また、変数増減法によって選択された歌唱中の音響・画像特徴量は嚥下障害に対して有効な予測をできないものの、構音障害に対してある程度有用な予測が可能であることが示された。この結果は、jitter, shimmer, HNR, dfa といった音響特徴量が歌唱中の音声にも有効である可能性を示唆している。

5. おわりに

本稿では、歌による嚥下・構音機能の定量的評価手法の

実現に向けて、歌唱データの音響・画像分析を行った。変数増減法による特徴量の選択の結果、構音障害は比較的分類性能が高い一方、嚥下障害では分類性能が低いことが確認された。今後の課題を以下に挙げる。

5.1 歌唱中の音節の正確な分割

今回の歌唱の音響分析では、楽譜上の音符ごとに区切られた音声を擬似的に音節とみなし、音響特徴量の計算を行った。しかし、図 4 に示した intial burst, vowel onset, occlusion などから計算される音響特徴量が、オーラルディアドコキネシスでの音響分析では構音障害の高い分類性能を発揮したことを踏まえると、歌唱中の音声のこれらの分析が分類性能の向上に寄与する可能性がある。前述の通り、オーラルディアドコキネシスに用いたアルゴリズムは、音節ごとの境界が曖昧になる歌唱には直接適用できないため、新たなアルゴリズムの考案を今後の課題としたい。

5.2 音響・画像分析に基づいた口腔機能の採点

歌唱を通じて分析された口腔機能を点数としてユーザーにフィードバックすることで、ユーザーの口腔機能の訓練のモチベーションの向上が期待できる。そのために、音響・画像特徴量を用いた口腔機能の採点方法について今後検討していきたい。

5.3 嚥下・構音障害のデータ収集

これまで我々が募集した参加者は、構音・嚥下機能に関して医師の診断を受けた人たちではなかったため、嚥下・構音障害の割合が少ないという問題がある。今後、我々は嚥下・構音機能に障害を抱える人たちのデータも収集していくことで、これを解決する予定である。

謝辞 本研究の実施や本論文の執筆に際し、多くのアドバイスをいただいた研究室の皆様へ深く感謝いたします。

参考文献

- [1] Ando, T., Masaki, A., Liu, Q., Ooka, T., Sakurai, S., Hirota, K. and Nojima, T.: Squachu: a training game to improve oral function via a non-contact tongue-mouth-motion detection system, *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*, pp. 1-8 (2018).
- [2] Belafsky, P. C., Mouadeb, D. A., Rees, C. J., Pryor, J. C., Postma, G. N., Allen, J. and Leonard, R. J.: Validity and reliability of the Eating Assessment Tool (EAT-10), *Annals of Otolaryngology, Rhinology & Laryngology*, Vol. 117, No. 12, pp. 919-924 (2008).
- [3] Belafsky, P. C., Mouadeb, D. A., Rees, C. J., Pryor, J. C., Postma, G. N., Allen, J. and Leonard, R. J.: Validity and reliability of the Eating Assessment Tool (EAT-10), *Annals of Otolaryngology, Rhinology & Laryngology*, Vol. 117, No. 12, pp. 919-924 (2008).
- [4] Dejonckere, P., Remacle, M., Fresnel-Elbaz, E., Woisard, V., Crevier, L. and Millet, B.: Reliability and

- clinical relevance of perceptual evaluation of pathological voices., *Revue de laryngologie-otologie-rhinologie*, Vol. 119, No. 4, pp. 247–248 (1998).
- [5] Fried, L. P., Tangen, C. M., Walston, J., Newman, A. B., Hirsch, C., Gottdiener, J., Seeman, T., Tracy, R., Kop, W. J., Burke, G. et al.: Frailty in older adults: evidence for a phenotype, *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, Vol. 56, No. 3, pp. M146–M157 (2001).
- [6] Fryzlewicz, P.: Wild binary segmentation for multiple change-point detection, *The Annals of Statistics*, Vol. 42, No. 6, pp. 2243–2281 (2014).
- [7] Godino-Llorente, J. I. and Goémez-Vilda, P.: Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors, *IEEE Transactions on Biomedical Engineering*, Vol. 51, No. 2, pp. 380–384 (2004).
- [8] Hu, K., Ivanov, P. C., Chen, Z., Carpena, P. and Stanley, H. E.: Effect of trends on detrended fluctuation analysis, *Physical Review E*, Vol. 64, No. 1, p. 011114 (2001).
- [9] Lee, C.-H., Liu, Z., Wu, L. and Luo, P.: MaskGAN: Towards Diverse and Interactive Facial Image Manipulation, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2020).
- [10] Lempel, A. and Ziv, J.: On the complexity of finite sequences, *IEEE Transactions on information theory*, Vol. 22, No. 1, pp. 75–81 (1976).
- [11] Little, M., McSharry, P., Roberts, S., Costello, D. and Moroz, I.: Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection, *Nature Precedings*, pp. 1–1 (2007).
- [12] Nishida, T., Yamabe, K., Ide, Y. and Honda, S.: Utility of the Eating Assessment Tool-10 (EAT-10) in evaluating self-reported dysphagia associated with oral frailty in Japanese community-dwelling older people, *The journal of nutrition, health & aging*, Vol. 24, No. 1, pp. 3–8 (2020).
- [13] Novotný, M., Ruzs, J., Čmejla, R. and Růžička, E.: Automatic evaluation of articulatory disorders in Parkinson's disease, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 22, No. 9, pp. 1366–1378 (2014).
- [14] Tanaka, T., Takahashi, K., Hirano, H., Kikutani, T., Watanabe, Y., Ohara, Y., Furuya, H., Tetsuo, T., Akishita, M. and Iijima, K.: Oral frailty as a risk factor for physical frailty and mortality in community-dwelling elderly, *The Journals of Gerontology: Series A*, Vol. 73, No. 12, pp. 1661–1667 (2018).
- [15] Viola, P. and Jones, M.: Rapid object detection using a boosted cascade of simple features, *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, Vol. 1, Ieee, pp. I–I (2001).
- [16] 伊藤加代子: 新しい口腔機能測定器を用いたオーラルディアドコキネシスの測定, *新潟歯学会雑誌*, Vol. 39, No. 1, pp. 61–63 (2009).
- [17] 小口和代, 才藤栄一, 水野雅康, 馬場尊, 奥井美枝, 鈴木美保: 機能的えん下障害スクリーニングテスト「反復唾液えん下テスト」(the Repetitive Saliva Swallowing Test: RSST) の検討 (1) 正常値の検討, *リハビリテーション医学*, Vol. 37, No. 6, pp. 375–382 (2000).
- [18] 西尾正輝, 新美成二: Dysarthria における音節の交互反復運動, *音声言語医学*, Vol. 43, No. 1, pp. 9–20 (オンライン), DOI: 10.5112/jjlp.43.9 (2002).
- [19] 日本医師会: オーラルフレイル対策のための口腔体操 | オーラルフレイル | 日本歯科医師会 (2021).
- [20] 公益社団法人日本歯科医師会: 歯科診療所におけるオーラルフレイル対応マニュアル 2019 年版 (2019).
- [21] 飯島勝矢: 食 (栄養) および口腔機能に着目した加齢症候群の概念の確立と介護予防 (虚弱化予防) から要介護状態に至る口腔機能支援等の包括的対策の構築および検証を目的とした調査研究, 事業実施報告書. 平成 26 年度厚生労働省老人保健事業推進等補助金. 老人保健健康増進事業 (2015).