情報処理学会論文誌 Vol.0 No.0 1–11 (??? 1959)

スマートフォンを用いた画像認識による嚥下機能の定量的評価手法

耿世嫻[1] 平井雄太[1] 下島銀士[1] 柳田陵介[2] 山田大志[2] 小野寺宏[1] 戸原玄[2] 矢谷浩司[1]

受付日 xxxx年0月xx日, 採録日 xxxx年0月xx日

概要：医師による嚥下障害の診断には，患者の定期的な通院が必要であり，診断結果は医師の経験に大きく影響されるという課題がある．そこで我々は，患者が在宅で簡単に嚥下障害の可能性の有無を評価できるように，スマートフォンで撮影した動画から嚥下機能を評価する手法を提案する．この提案手法の実現に向けて，必要なタスク群を決定し，147 名の実験参加者から得られた動画を分析した．その結果，70.6% の精度（Balanced Accuracy）と Weighted F1 score は 0.801 でランダムの予測スコアより 0.243 ポイント高い識別性能を得ることができた．

キーワード：モバイルヘルス，高齢化と健康，嚥下困難，嚥下障害，画像認識

A Vision-based Oral and Swallowing Capability Quantification Approach with Smartphones

SHIXIAN GENG[1] YUTA HIRAI[1] GINSHI SHIMOJIMA[1] RYOSUKE YANAGIDA[2] TAISHI YAMADA[2] HIROSHI ONODERA[1] HARUKA TOHARA[2] KOJI YATANI[1]

Received: xx xx, xxxx, Accepted: xx xx, xxxx

Abstract: The diagnosis of dysphagia by clinicians poses challenges, as it requires patients to make regular clinic visits and the diagnosis requires training and empirical skills of the doctors. To address these issues and enable patients to assess potential symptoms of dysphagia easily at home, we propose a method to evaluate oral and swallowing function from videos captured by smartphones. To achieve this, we determined the required set of tasks for analyzing oral and swallowing function and analyzed videos of these tasks obtained from 147 participants. As a result, we obtained a discrimination performance with a balanced accuracy of 70.6% and a Weighted F1 score of 0.801. The Weighted F1 score of our model (0.801) is 0.243 points better than that of a random predictor.

Keywords: Mobile health, aging and health, swallowing difficulty, dysphagia, image recognition

1. はじめに

嚥下障害とも呼ばれる，上手く飲み込みができない障害（dysphagia）は，高齢者の間で頻繁に見られ，咳や窒息，栄養不足など多くの健康上の問題を引き起こす可能性がある [27]．アメリカの統計報告によると，2014 年には 1 年間に 25 人に 1 人が嚥下の問題を抱えており，そのうち約 3 分の 1 の患者がこれを大きな問題と捉え，しかし嚥下の問題を診断されたのはそのうち半数未満と報告されている [6]．嚥下障害は，脳卒中発症後の症状の一つとして現れることがあり，パーキンソン病，口腔がん，または加齢に伴う顔面筋力低下などの病気によっても引き起こされることがある [27]．症状が現れた場合，現在の評価方法では，通常，歯科医師の経験を要する検査が必要とされるため，患者は頻繁に病院を訪れて状態を追跡する必要がある．しかし，コロナウイルス感染症や物理的制約などにより，対面での診察が難しいことがある．さらに，検査は主に医師の経験によるものであり，患者は医師から質的なコメントしか受け取ることができないため，血圧や肺機能検査などの健康

[1] 東京大学
[2] 東京医科歯科大学 摂食嚥下リハビリテーション学分野

診断と異なり数値的な測定データを得ることが難しい．歯科病院では舌圧測定器や口唇閉鎖力測定 [14,16,38] などの医療機器が使用されることがあるが，これらの情報は嚥下障害の評価には限界があるため，患者が自宅で使用することはできず医師の指導が必要とされている．このため，現在の医療環境において，嚥下障害の患者が自分の状態を継続的に追跡することは困難である．

嚥下障害を定量化し，診断を支援する方法の 1 つは，Iowa Oral Performance Instrument（IOPI）[14] のような医療用舌圧測定器を使用することである．この医療機器は，使い捨てバルーンとプローブで構成されている．デバイスを使用するには，患者がバルーンを口に入れ，舌で可能な限り力を加える必要がある．バルーンが変形し中の空気を押し出すことによって，デバイスはその圧力の変化を測定し，舌によって加えられた圧力を推定することができる．この器具は口に入れ舌に密着させる必要があるため，医師の指導が必要である．

専門家との議論によると，このようなデバイスは縦方向の舌運動に関連する垂直な舌力を測定するように設計されている．しかし，嚥下障害の場合，横方向の舌運動が縦方向の運動よりも重要であり，現状のデバイスの設計では横方向の舌力を測定することは難しい．

また，これらのデバイスを使用して舌圧または唇の力を測定することは，プローブを患者の口の中に置く必要があるため，比較的侵襲的であると考えられる．さらに，このようなデバイスは個人での購入は想定されておらず，比較的高価であることが問題点になる．

オーラルディアドコキネシス [1] は，嚥下障害および構音障害の評価の 1 つで，一般的に使用されている他の方法よりも簡単であり，より定量的なものである．これは，患者が 10 秒間に「pa」，「ta」，「ka」という音節を何回繰り返すことができるかを数えることで，嚥下障害を測定するものである．飲み込みの問題を抱えた患者は，これらの音節をしっかりと繰り返すことが困難である傾向がある．信号処理技術を用いて，自動的にオーラルディアドコキネシスのカウントを行う医療機器 [15] やスマートフォンアプリケーション [10] も存在している．オーラルディアドコキネシスは単純で直感的ではあるが，発音を数えることだけでは飲み込み能力を追跡するには不十分である．

もう 1 つの単純で効果的なテストは，医療機器を必要としない，反復唾液嚥下テスト（RSST）である [20]． RSST では，患者に自分自身の唾液を 30 秒以内にできるだけ飲み込むように指示する．このときの飲み込み行動が何回起こったかを数えることで，飲み込み障害の評価ができる．しかし，患者が頻繁にこのプロセスを実行することは，飲み込む際に痛みや困難を覚えることがあるため，やる気を喪失することがある．

本研究の目的は，現状の通院に加えて，嚥下能力を継続的に追跡するのを支援できる非侵襲的な方法を開発し，量的な測定で障害に関するより多くの情報を提供することである．特にスマートフォンを利用したセンシングによる，嚥下機能の測定を実現することを目的とする．我々の研究は，嚥下障害患者がスマートフォンのカメラを使って自宅で簡単に自身の状態を知ることができるシステムに関して初めての研究であり，概念実証として位置づけられる．

2. 医学的背景

嚥下が困難になるにはさまざまな原因がある．一般的な原因の 1 つは，嚥下の過程に重要な神経の損傷である．脳卒中やパーキンソン病などの神経学的障害は，嚥下過程を制御する神経を損傷する可能性がある．正常な嚥下には，顔の筋肉，舌，唇，喉などの様々な部位が正しく機能し，協力する必要がある．これらの部位を制御するために，いくつかの神経がその役割を担っている．ここでは嚥下において重要な神経を示す [7]：

- 顔面神経（脳神経 Ⅶ）：顔面神経（脳神経 Ⅶ）は，目を閉じる，口を突き出す，歯を見せるなどの顔の動きを制御する神経である．嚥下過程には，顔面神経も運動を仲介している．顔面神経に関連する障害は，顔の一部の麻痺や非対称，または無意識の動きを引き起こすことがある．
- 舌咽神経（脳神経 Ⅸ, Ⅴ）：舌咽神経と迷走神経とも呼ばれ，嚥下時に咽頭を活動させる神経である．また，発声にも関与している．舌咽神経と迷走神経の障害は，嚥下障害や発語障害を引き起こすことがある．
- 舌下神経（脳神経 Ⅻ）：舌下神経とも呼ばれる脳神経 Ⅻ は，舌の下を通り，嚥下を容易にするための舌の運動を制御する神経である．舌下神経の障害は，異常な舌の動きを引き起こすことがある．

異なる脳神経の中で，脳神経 Ⅶ や脳神経 Ⅻ の障害は，スマートフォンのカメラで撮影した動画で評価できる程，比較的明らかな症状を引き起こす．脳神経 Ⅶ の損傷はしばしば顔の動きの異常を引き起こし，脳神経 Ⅻ の損傷は舌の動きを困難にすることがある．脳神経 Ⅸ/Ⅴ の損傷をスマートフォンのカメラで評価するのは難しく，喉頭の接近検査が必要である．そのため，本研究では，脳神経 Ⅶ と脳神経 Ⅻ の障害による嚥下障害を，神経学的障害として考慮した．

神経学的障害以外に，嚥下障害の原因として，加齢に伴う筋肉の衰弱や口腔がんなどがある [27]．原因には違いがあるにもかかわらず，しばしば似たような結果につながることがある．患者は，舌を突き出す，頬を膨らませるなどの口腔運動を健康な人と同じように行うことが困難であることが多い．しかし，実際に適切に嚥下を行うためには

ある種の口腔運動が必要となる．したがって，口腔運動の異常を検査することにより，嚥下困難を評価することができる．

日本では，医師や歯科医師は，評価基準書である「標準ディサースリア検査」（AMSD）[36] や「嚥下運動機能検査」（AMFD）[37] に従って診断を行うことが多い．これらは，患者に様々な条件（座位/臥位）下で，顔・口の動きや発声機能，呼吸，液体の飲み込みなど，20 以上のタスクを実行してもらい，嚥下機能を評価する．したがって，チェックには数時間かかる可能性がある．訓練された専門家は，患者がこれらのタスクをどの程度達成できるかを観察し，経験に基づいて定性的な判断を下す．そのため，決定がどのように行われるかは，タスクの複雑さや経験的な性質のため，しばしば説明できない状況である．このような評価を行うには，関連する医学的な背景を持ち，綿密な訓練を行なった医師が必要である．また，高齢患者にとっては，このようなテストは負担がかかることがある．実際の生活では，患者は頻繁に病院を訪問することや長時間テストを受けることに抵抗があることがある．

AMSD や AMFD のようなテストは，患者の状態に関する包括的な情報を提供できるが，自己評価のために機械学習アルゴリズムを開発して完全な評価手順を再現することは不可能であり望ましくない．したがって，私たちの目標は，少なくかつ簡単なタスクで構成され，通院の診察と組み合わせて使用できるシステムを構築し，患者がスマートフォンを使って自宅で嚥下機能の状態を追跡し理解するための付加的で定量的な情報を提供することである．

3. 関連研究

嚥下機能は，言語聴覚療法でトレーニングできる．言語聴覚療法では，異なる方向に舌を突き出したり，異なる言葉を発声したりするなど，さまざまな口腔の運動を行う．トレーニングを行わない場合，嚥下障害の状態は悪化する可能性があるが，良好なトレーニングによって症状を改善させることができる．したがって，嚥下障害のある人々は，自分の状態を毎日継続的に追跡し，自身の状態が悪化していないかどうかを把握し，適切な治療とトレーニングを受ける必要がある．企業や研究者は，嚥下障害の継続的な追跡を促すために，さまざまな方法を開発してきた．

3.1 音響による嚥下機能の評価

先行研究では，首に装着可能なセンサーを用いて嚥下の音を感知する装置が提案された [9]．また，PLIMES 社は，「Gokuri」という商用プロトタイプの製品化プロジェクトを始動させている [22]．この商用製品によって感知される嚥下音に関する情報は，無線通信を介してスマートフォン上に表示されるように設計されている．しかしながら，首は人体の非常に重要な構造であるにもかかわらず，このような実装によって，患者の首に密着することになるため，安全上の懸念が生じる可能性がある．このような製品には，会社が独自に開発されたハードウェアが含まれているため，手に入れるには高価になる場合もある．

音響による嚥下及び構音の評価も幅広く探究されている．研究者たちは，患者の声を分析して声の障害を検出するアルゴリズムを開発しており，これは嚥下困難に関連する可能性がある [11, 12]．しかし，この手順は単調であり，患者が興味を持って継続的に行うことが難しい場合がある．オーラルディアドコキネシスもまた，非常に単調な作業であることがある．我々の以前の研究 [35] では，カラオケを用いたゲーム化によって音響センシングによる構音や嚥下困難の継続的なモニタリングを促進する可能性を探究していた．

3.2 音響センシング以外の嚥下機能評価

また，嚥下機能の評価には画像認識を用いた方法も研究されている．例えば，画像認識を用いてサルコペニアによる嚥下障害を評価し，頸部の静止画像の画像処理に注目した研究がある [24]．しかし，嚥下機能の低下は，喉の機能低下だけでなく，口を開けることができない，舌を出すことができないなどの症状にもつながることがある．これらの器官も嚥下の過程で重要な役割を果たし，また，これらの器官の動きも評価時の重要な特徴である．したがって，画像データだけでなく，動画データも充実させ，自宅で簡単に実施できる評価方法を探求する必要があると考えている．

4. データ収集

現在の嚥下機能の評価方法は主に経験に基づくものが多く，実験全体を設計する際には，医師の意見を十分に考慮する必要がある．そこで，本研究では，神経内科医数名に何度かインタビューを行い，共同で実験およびデータ収集の方法について議論を行った．

4.1 医師インタビュー

システムの最も重要な特徴の 1 つは，嚥下能力の頻繁でかつ継続的な追跡を促進するため，利用が容易であることである．そのため，複雑な指示や長時間のテストは好ましくない．評価を行うために十分な特徴を得ることができるように，私たちは東京大学病院の医師や，嚥下に関連する医療背景を強く持つ神経内科・歯科の医師と数回のインタビューを行い，彼らの経験や AMSD [36]， AMFD [37] といった臨床で用いられる手法に基づいて，システムとデータ収集実験手順を共同で設計した．

前々章で述べた通り，嚥下過程では舌機能と顔面筋が重

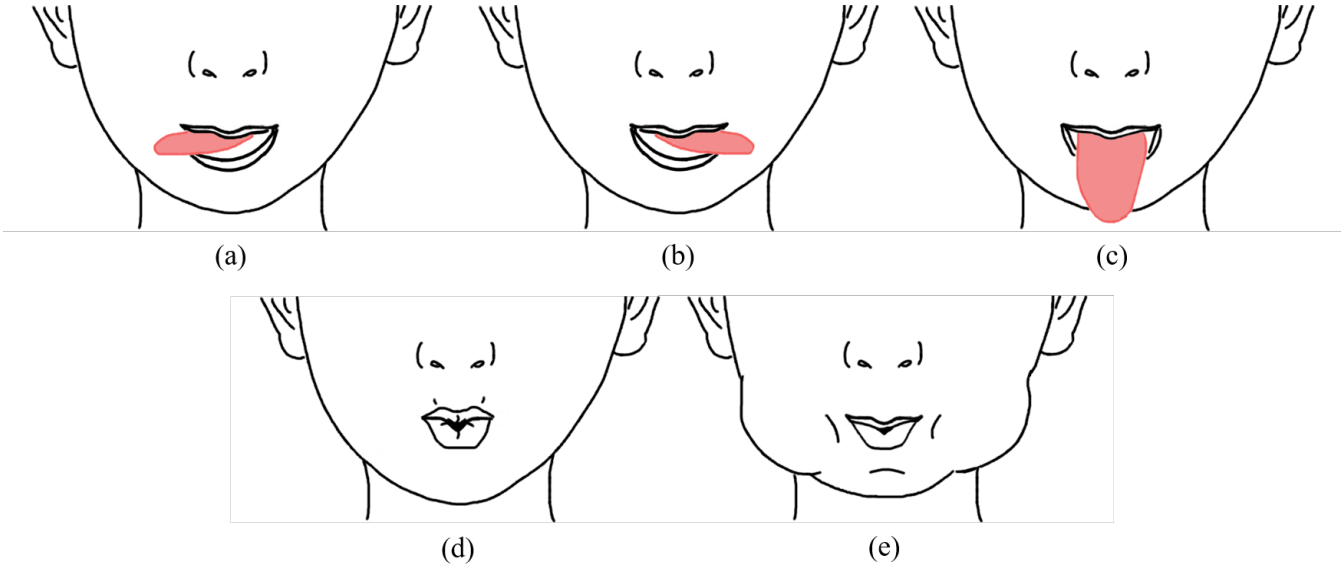


図 1: 撮影した口・舌の動きの種類．(a) 左に舌を突き出す (b) 右に舌を突き出す (c) 前に舌を突き出す (d) 口をすぼめる (e) 頬を膨らませる．

要であるが，それらを評価するため，議論の結果，以下の 5 つのジェスチャーを最終的に決定した．喉は嚥下能力にとっても重要な構成要素であるが，スマートフォンのカメラで喉内部の詳細を捉えることは非常に困難であるため，喉の状態を評価するためのタスクは含めなかった．また，プライバシーを保護するため，参加者の顔の下半分のみを撮影した．図 1 に対応する 5 つの口腔タスクを示している．

4.2　実験参加者の募集

本研究では，以下の 3 つの参加者グループから参加者を募集した：1）クラウドソーシング（Crowdworks）：20 歳から 60 歳までの参加者．2）シルバー人材センター：60 歳から 90 歳までの高齢者．3）嚥下障害と診断された 52 歳から 92 歳までの患者．

グループ 1 の参加者のビデオは，実験室ではなく異なる環境条件下で，様々な機種のスマートフォンを用いて，参加者自身で撮影してもらった．グループ 2 およびグループ 3 の参加者のデータは，それぞれ異なる日付にて同じ部屋で収集された．グループ 2 では撮影機材のセットアップを使用したが，参加者が椅子を動かすことができるためカメラとの距離は固定されていなかった．カメラ自体は三脚に固定されていたが，参加者の頭の動きによって，他の 2 つのグループと同様の結果が得られることがわかった．グループ 3 では，データを収集する医師に撮影機材を提供したが，実際にデータを収集する際には使用されなかった．その結果，3 つのグループともカメラとの距離が固定されておらず，カメラからの視野も安定していない，似たような条件であったと考えられる．

我々は収集したデータに検証作業を行った．ビデオの画像が暗く，顔がはっきりと見えない場合は，データを除外した．また，タスクの維持時間が非常に短い，舌や唇，鼻がフレーム内に完全に捕捉されていない，解像度が非常に悪いなど，最低基準を満たさないデータも除外した．さらに，現代のスマートフォンのカメラに最も広く使用されている 30fps のフレームレートで収録されたビデオのみを分析に採用した．

さらに，参加者にはアンケートに回答してもらった．内容は年齢や生物学的性別など関連する人口統計に関する質問に加え，嚥下障害のための広く使用されている自己評価指標である Eating Assessment Tool（EAT-10）[5] を含む．EAT-10 は 3 点以上のスコアを取得した場合，嚥下障害の可能性があるとされ，医師や歯科医師による正式な評価を受けることが推奨されている．本研究では，参加者がどのグループに属しているかに関わらず，スコアが 3 点以上の参加者を嚥下障害の可能性があるとラベルを付けた．この情報は，機械学習アルゴリズムの訓練のための正解データとして後に使用される．EAT-10 スコアに基づくラベリング基準については，歯科医師と相談し，妥当であると認められている．

合計で，先述の 3 つのグループから 147 人の参加者から，不適格なデータを除いたデータを収集した．すべての参加者の中で，26 人が EAT-10 で 3 以上のスコアを獲得し，嚥下障害の可能性があるとラベル付けされた．これらのうち，14 人は 3 番目のグループ，つまり病院から来た参加者であり，残りの 12 人は他の 2 つのグループに属する参加者である．なお，3 つのグループのうち，グループ 3 の参加者のみが病院を受診し，医師から正式な診断を受けていた．詳細は表 1 に示されている．

表 1: 参加者情報

	参加者数	グループ概要
グループ 1	74（うち 5 名が EAT-10 スコア 3 点以上）	クラウドソーシングで募集, 若者, 健康
グループ 2	58（うち 7 名が EAT-10 スコア 3 点以上）	シルバー人材センター経由で募集, 高齢者, 健康
グループ 3	15（うち 14 名が EAT-10 スコア 3 点以上）	病院, 患者

5. データ分析

口のジェスチャー映像から 24 個の特徴を抽出した．そのうち 14 個は舌と唇の長さと角度から計算され，10 個は舌と唇の振動を表している．それらの特徴量と説明は表 3 に示している．舌と唇の長さと角度，特徴量でいうと#1〜14 は，2 節で述べた舌下神経や顔面神経を評価する指標となるため，AMSD [36] や AMFD [37] では医師がよく検査する項目である．例えば，舌下神経を診るために，舌をどれだけ長い時間いろいろな方向に出すことができるか，また，顔面神経を診るために，どれだけ顔を動かすことができるかによって，医師は患者に点数をつける．舌と唇の振動，特徴量でいうと#15〜24 は，本研究第六著者がパーキンソン病などの神経疾患は嚥下障害につながる可能性があり，その指標の一つとして表情筋や舌の不随意運動が考えられるとしたことから，追加で調査したものである．

5.1 長さと角度

実験参加者がジェスチャーを維持しようとしているフレームの中からランダムに 1 つを抽出し，舌の突き出しの長さを，舌先の座標と上唇の中央の座標の距離としてピクセルで計算した．当初，既存の顔ランドマーク検出器 [8,17,32] の使用を考えていたが，下半分の顔のデータではうまく機能せず，また，本研究では顔の下半分に対するランドマーク検出器の開発や微調整は主要な目的ではないため，手動で舌先と上唇の中央をラベル付けした．このピクセル値を実際の長さに変換するために，AIST 人体寸法データベースの日本人の平均鼻幅を用いた [2]．表 2 に，鼻幅の平均値と標準偏差を示す．同じ理由で，実験参加者の鼻の左右の点を手動でラベル付けした．解析コードは，鼻幅をピクセル単位で計算し，次の式を使用して，舌の長さをミリメートル単位で求めた：$L_{\text{target}} = L_{\text{nose}}/P_{\text{nose}} * P_{\text{target}}$．ここで，$P_{\text{nose}}$，$L_{\text{nose}}$，$P_{\text{target}}$ は，それぞれ，収集したビデオデータの鼻幅（ピクセル），AIST データベースによる平均鼻幅，および計測された舌の長さ（ピクセル）である．口をすぼめる，頬を膨らませるジェスチャーにおいては，唇の水平長さおよび唇の垂直長さを測定し，また唇の向いている角度を測定した．COVID-19 による影響を考慮して，参加者の鼻幅の推定に AIST のデータベースを用いることにした．本研究の最も重要な対象者である高齢者はウイルスに感染しやすいため，データ収集の際には身体的接触を避けて適切な社会的距離を保ち，実際の鼻幅の測定は行わなかった．表 2 に示すように，各項目の標準偏差はそれほど大きくなく，平均的な鼻の広さを基準としても比較的信頼できることを示している．さらにこのデータベースには，サンプルで測定された鼻幅の最大値と最小値も含まれており，これらの値は平均値から約 15 %ずれている．つまり鼻幅の誤差は，最悪の場合で我々の計算した長さに約 15 %の誤差をもたらす可能性がある．AIST のデータの標準偏差が小さいことからこのようなケースは比較的まれであると考えられるが，概念実証となる本研究としてはこの誤差は許容範囲であると考える．

表 2: AIST データベースによる鼻幅平均値（標準偏差） [2]

	若年層	年長者
男性	36.2 (2.40) mm	38.7 (2.52) mm
女性	33.1 (1.88) mm	35.9 (2.62) mm

5.2 動き

我々は舌や顔の神経の損傷がジェスチャーを不安定にする可能性があるため，口や顔の震えも特徴量に含めた．この病態は，顔，顎，舌の無意識かつ異常な収縮で顎口腔ジストニア（OMD）と呼ばれる [21]．その理由は，AMSD または AMFD の嚥下の臨床的評価では通常考慮されないが，これらの特徴量がどの程度説明的であるかについて知りたかったためである．

本研究では，Kanade-Lucas-Tomasi（KLT）[26] を用いてこの動きを追跡した．最初のフレームで口や舌のキーポイントを手動でラベリングした後，それらのキーポイントを KLT で自動的に追跡することで，口や舌の動きの特徴量を抽出した．また，撮影する時に手やスマートフォンの動きなどの全体的な動きの影響を除くために，鼻の端のキーポイントからの相対座標として考えた．

ジェスチャーを安定して保持しているフレーム（表 3 の下半分）における，各タスクの平均特徴量は，以下の式で算出される．

$$m_{\text{avg}} = \sum_{n=f_{\text{start}}}^{f_{\text{end}}-1} \sum_{i=1}^{5} \frac{F\sqrt{(x_{i,n} - x_{i,n+1})^2 + (y_{i,n} - y_{i,n+1})^2}}{5(f_{\text{end}} - f_{\text{start}})} \tag{1}$$

ここで f_{start} 及び f_{end} は，各タスクの動画データにおいて，タスクを開始及び終了したフレーム番号を表す．i は表

3 中に示した 5 つのラベル付き点のインデックスである．

6. 結果

6.1 機械学習

24 個の特徴量の計算後，各特徴量を，放射状基底関数 （Radial Basis Function, RBF）をカーネルとするサポートベクターマシン （Support Vector Machine, SVM）への入力とした．

我々のデータセットにおいて EAT-10 スコアより嚥下障害の可能性があるものは 26 例しか存在しなかったため，本研究では，カーネルスケールやボックス制約などのモデルパラメータの訓練および調整を，訓練データとテストデータを 7:3 の割合で分割したデータを用いて行った．その後，過学習を避けるため，モデルの評価にあたり，5 重の交差検証を実施した．これは，各グループに約 30 のサンプルがあり，そのうち，我々が適切と考えた割合，20 % がテストに使用されることを意味する．モデルの性能は，ROC（Receiver Operating Characteristic）曲線および PR（Precision-Recall）曲線をプロットし，評価した．また，PR 曲線に基づいて分類性能が最適化されるよう予測閾値を調整したところ，再現度は 0.577，適合度は 0.429 となった．

不均衡なデータセットでは，ROC 曲線（図 2）は偏ることがあるが，PR 曲線（図 3）は偏らない．図に示すように，我々のモデルは妥当な ROC 曲線（AUC = 0.6745，PR 曲線（AUC = 0.3840）を示した．この結果は非常に優れているとは言いがたいが，得られた F_1 値（0.492）はランダムの予測スコア（0.261）より 0.231 ポイント高い．PR 曲線では，再現度が 0.4 と 0.6 あたりで精度が急激に低下し，モデルがいくつかの正負のケースを区別することが困難であることを示している．

我々のデータセットが不均衡であることも踏まえ，最終的なモデルの精度（Balanced Accuracy）は 70.6%であり，Weighted F1 score（クラスごとに F_1 値を計算してデータ数で加重平均をとった値）は 0.801 であった．表 4 は 5 重交差検証による予測結果混同行列である．この結果は，ランダムで予測する Weighted F1 score（0.558）より 0.243 ポイント高い．

機械学習による予測は EAT-10 のスコアが 3 点以上かどうかであるが，その結果，医師から嚥下障害陽性と診断されている 15 名中 12 名を正確に予測できた．これは，比較的に深刻な嚥下障害陽性のケースに対しては、高い分類精度を示している．

6.2 相関分析

また，データの背後にある医学的な意味を見出すために，特徴量について Spearman の相関分析を行った．これまでの研究によって EAT-10 のスコアが 3 点以上の場合は嚥下機能に障害がある確率は高くなるものの，4 割以上は問題がないことが示唆されている [33]．したがって，すでに医師によって嚥下障害陽性例と確認されたグループ 3 の参加者のみを扱うほうがより正確である可能性がある．SVM による分類では，学習データの陽性サンプル数が著しく減少するためこの方法を用いることはできなかったが，相関分析では適用することができる．さらに，我々のデータは特徴量と嚥下障害の間に単調関係があると仮定でき，かつ各観測が独立であり，変数は少なくとも順序変数であることから，Spearman の相関分析の仮定を満たしている．図 4 は，Spearman の相関係数とその p 値を示したものである．このプロットから，左・右・前に舌を突き出したときの舌の長さと口をすぼませた・頬を膨らませたときの唇の長さは嚥下障害陽性例（グループ 3: 病院患者）と負の相関を持つことがわかる．このことから，嚥下障害陽性例では，舌の突出長が比較的小さく，また，タスク 4 とタスク 5 では，唇の垂直方向についてその長さが小さいことから，これらのタスクを実行できない可能性があることが示された．一方，口をすぼませる・頬を膨らませる唇の動きは嚥下障害と正の相関があり，タスク 4 と 5 で保持時の口また顔の震えが嚥下障害と相関している可能性があることが示唆される．

7. 本結果を用いた想定されるシステム

本研究の成果を活用し，モバイルアプリを用いたシステムを設計する予定である．このシステムでは，ユーザからデータを取得した後，ソフトウェアプロトタイプの前処理技術として顔のランドマーク検出器を使用することで，本実験と同様に，ユーザがキーポイントにラベル付けする必要がないようにする予定である．提案システムの全体的なユーザシナリオを，図 5 に示す．最初に，ユーザは指示に従い，スマートフォンのフロントカメラで 5 つの口腔内を動かすタスクを撮影する．また，本研究から殆どの参加者が 5 つのタスクを 3〜10 分以内に終えられることもわかっている．次に，顔のランドマーク検出器により特徴量抽出を行う．そして，その結果を機械学習アルゴリズムに入力し，モデルの予測に基づき，評価レポートを作成する．ユーザが自分でキーポイントをラベリングすることは面倒な場合があるため，一般的な顔のランドマーク検出器をスマートフォン上で使用する [4,13]．これにより，少なくとも鼻，唇，頬を検出することができる．また，これらのキーポイントのみをクラウドに送信して解析することで，ユーザのプライバシーを保護することができる．ただし，舌の検出に関しては，まだ十分に開発されたモデルがないため，可能であれば，舌のキーポイントを追跡するための舌検出アルゴリズムを独自に構築したいと考えている．それができなかった場合，ユーザが手動で舌のキーポイントにラベル

表 3: タスクの一覧

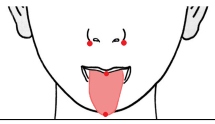
	キーポイント（赤丸の部分）	特徴量
タスク 1		左に舌を最も突き出したときの 舌の長さ（特徴量 #1）， 顔の垂直中心線との角度（特徴量 #2）
タスク 2		右に舌を最も突き出したときの 舌の長さ（特徴量 #3）， 顔の垂直中心線との角度（特徴量 #4）
タスク 3		前に舌を最も突き出したときの 舌の長さ（特徴量 #5）， 顔の垂直中心線との角度（特徴量 #6）
タスク 4		口をすぼめて安定したときの 水平・垂直方向の唇の長さ（特徴量 #7 ， #8）， 顔の水平中心線に対する水平・垂直方向の唇の角度（特徴量 #9， #10）
タスク 5		頬を膨らませて安定したときの 水平・垂直方向の唇の長さ（特徴量 #11 ，#12）， 顔の水平中心線に対する水平・垂直方向の唇の角度（特徴量 #13 ，#14）
タスク 1		保持時の舌上 5 点の 平均的な動き（特徴量 #15）， 動きの標準偏差（特徴量 #20）
タスク 2		保持時の舌上 5 点の 平均的な動き（特徴量 #16）， 動きの標準偏差（特徴量 #21）
タスク 3		保持時の舌上 5 点の 平均的な動き（特徴量 #17）， 動きの標準偏差（特徴量 #22）
タスク 4		保持時の唇上 5 点の 平均的な動き（特徴量 #18）， 動きの標準偏差（特徴量 #23）
タスク 5		保持時の唇上 5 点の 平均的な動き（特徴量 #19）， 動きの標準偏差（特徴量 #24）

表 4: 混同行列

	予測 N	予測 P
実際 N	**101**	20
実際 P	11	**15**

を付ける必要がある．複雑な臨床評価であれば数時間かかることもある [36,37] が，ラベリングと分析プロセスの自動化によって，患者はこの手順を数分以内に素早く終えることができる．

また，本研究はスマートフォンを用いたカメラ映像に基づく簡便な嚥下能力評価の概念実証であるため，照明条件が悪いなどの不適格なデータは手作業で除外した．実際のシステムを考えたとき，そのインターフェースはより良いデータを取得するために改善することができる．例えば，

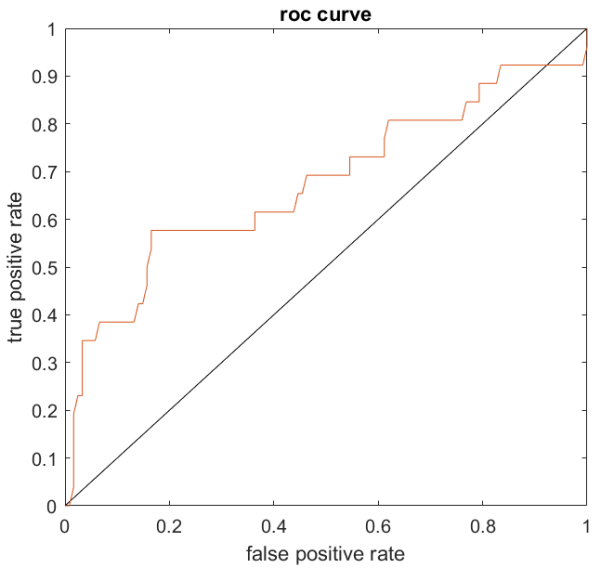


図 2: ROC 曲線（AUC = 0.6745）．AUC ががベースラインの 0.5 を超え，ランダムなモデルよりも優れた分類性能を持つと考えられるが，データセットのクラス分布が不均衡な場合，ROC 曲線は偏ることがある．

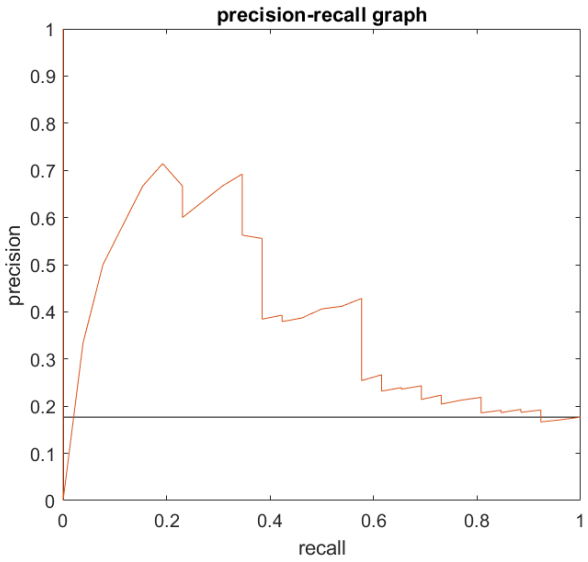


図 3: PR 曲線（AUC = 0.3840）．AUC がベースラインの 0.177 を上回っているため，一定程度の正例と負例を分類していると推察される．また，不均衡なデータセットでも，RP 曲線は偏らないので，ROC 曲線より信憑性がある．

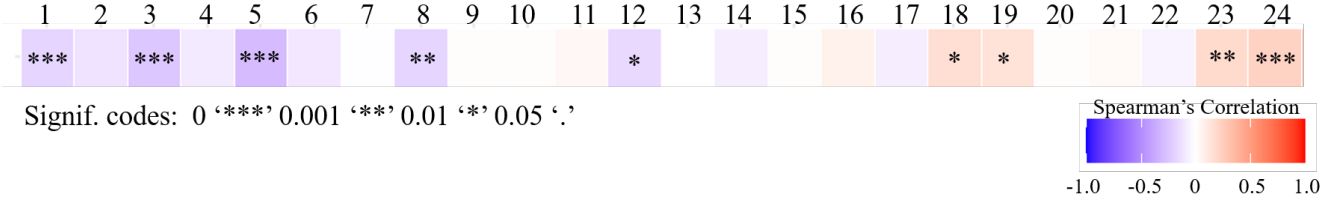


図 4: Spearman の相関分析: 左・右・前に舌を突き出したときの舌の長さと口をすぼませる・頬を膨らませるときの唇の長さは病院患者例と負の相関を持つことがわかる．一方，口をすぼませる・頬を膨らませるときの唇の動きは正の相関がある．

照明状態を検出するためにスマートフォンの輝度センサーを使用したり，顔全体がカメラフレーム内にあるかどうかをチェックするために顔のランドマーク検出器を使用したりすることが挙げられる．このような機能を統合することでデータの質は大幅に向上すると考える．

最後に，システムをゲーム化することで，ユーザが長期に渡ってシステムを継続的に使用する動機付けを与えることができる．例えば舌の動きで操作するビデオゲームを設計した先行研究がある [3] が，同様のインタラクションをスマートフォン上のシステムに統合することが考えられる．

8. まとめと今後の課題

嚥下機能の低下，例えば嚥下障害や失語症などは，身体能力に対してより深刻な負の影響を与える可能性がある．本研究では，スマートフォンのカメラで撮影された動画を用いて嚥下機能を推定することの実現可能性を検討した．嚥下障害の可能性有無の分類において，Weighted F1 score は 0.801 である．ランダムなモデルよりある程度良い性能であるが，これは今後の研究でさらなる改善を目指す．

本研究にはいくつか課題がある．まず，我々のデータセットには，限られた数しか構音障害や嚥下障害の症例が含まれていない．これにより，特徴量の選択にバイアスが生じる可能性がある．嚥下障害は非常に複雑な疾患であり，医師は慎重に検討する必要がある．複雑な評価方法を使用することで，より信頼性の高い結果が得られる可能性があるが，私たちのアプローチは医師による嚥下障害の臨床評価を代替するものではない．また，今回の研究では脳神経経 IX と V の評価は含まれていない．嚥下障害の原因として考えられるものはすべて医師によって検討される必要がある．第二に，我々のデータは 2 次元の動画でのみ構成されているが，実際には画像の深度マップは動きや長さを計測するために役立つため，これを考慮することが重要である．最新のスマートフォンモデルには専用の深度カメラが装備されているものもあるが，多くのスマートフォンには

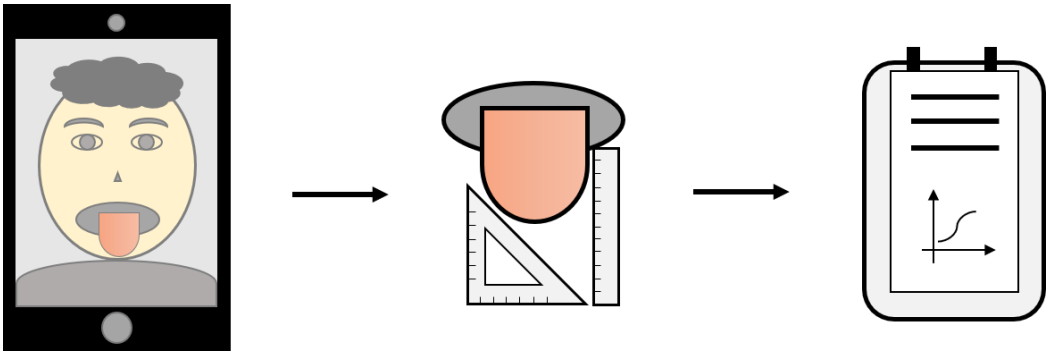


図 5: 本研究の実験結果を用いて構築することを想定しているシステム．ユーザはまず，スマートフォンのフロントカメラで決められた手順で動画を撮影する．その後，システムはさまざまな特徴量を定量化する．その特徴量を SVM に入力し，予測結果に基づいて評価レポートを作成する．

ついてないため，本研究では深度マップを考慮しないことにした．さらに，我々は EAT-10 のスコアを使って，データを訓練する際の正解ラベルを付与している．しかし，これまでの研究によって EAT-10 のスコアが 3 点以上の場合は嚥下機能に障害がある確率は高くなるものの，4 割以上は問題がないことが示唆されている [33]．したがって，すでに医師によって嚥下障害陽性と確認されている，病院から募集したグループ 3 の患者のみを嚥下障害陽性としてラベル付けすれば，機械学習モデルの分類性能がより良くなることを期待できる可能性がある．しかし本研究ではグループ 3 の嚥下障害陽性例は 14 例と限られていたため，グループ 1 と 2 に属する EAT-10 のスコア 3 点以上のデータも含めることにした．本研究では，確認された嚥下障害陽性例のサンプル数が比較的少なかったため，今後の研究ではより多くの嚥下障害陽性例を対象とした大規模な研究が必要である．これにより，機械学習モデルの性能が向上するだけでなく，医学的意義を見出すためにより多くの変数の相互作用を伴う多重ロジスティック回帰などの統計分析を行うことができるようになる．

謝辞 システム設計やデータ収集の過程でアドバイスをいただいた浜松医科大学の長島優先生に厚く御礼申し上げます．

参考文献

[1] Ackermann, H., Hertrich, I., and Hehr, T.: Oral diadochokinesis in neurological dysarthrias. Folia Phoniatr Logop, Vol.47, No.1, pp.15–23 (1995).

[2] AIST 人体寸法データベース 1991-92, AIST Digital Human Laboratory (オンライン), 入手先 (https://www.airc.aist.go.jp/dhrt/91-92/) (参照 2021-09-02).

[3] 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. In Proceedings of the 2018 International Conference on Advanced Visual Interfaces, pp. 1-8 (2018).

[4] Tracking the User's Face in Real Time, Apple (online), available from (https://developer.apple.com/) (accessed 2021-09-02).

[5] Belafsky, P.C., Mouadeb, D.A., Rees, C.J., et al.: Validity and reliability of the Eating Assessment Tool(EAT-10). Annals of Otology, Rhinology & Laryngology, Vol.117, No.12, pp.919-924 (2008).

[6] Bhattacharyya, N.: The prevalence of dysphagia among adults in the United States. Otolaryngology–Head and Neck Surgery Vol.151, No.5, pp.765–769 (2014).

[7] Brazis, P., Biller, J., and Gruener, G.: DeMyer's The Neurologic Examination: A Programmed Text (Seventh Edition), McGraw-Hill Education (2017).

[8] Chandran, P., Bradley, D., Gross, M., et al.: Attention-Driven Cropping for Very High Resolution Facial Landmark Detection. IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). pp.5860–5869 (2020).

[9] Choi, Y., Kim, M., Lee, B., et al.: Development of an Ultrasonic Doppler Sensor-Based Swallowing Monitoring and Assessment System. Sensors, Vol.20, No.16, p.4529 (2020).

[10] DDK: Diadochokinetic Assess, Collin Dunphy at App Store (online), available from (https://apps.apple.com/tr/app/ddk-diadochokinetic-assess/id1489873060) (accessed 2021-09-02).

[11] Godino-Llorente, J.I. and Gómez-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).

[12] Godino-Llorente, J.I., Gomez-Vilda, P., and Blanco-Velasco, M.: Dimensionality reduction of a pathological voice quality assessment system based on Gaussian mixture models and short-term cepstral parameters. IEEE Transactions on Biomedical Engineering, Vol.53, No.10, pp.1943-1953 (2006).

[13] Detect faces with ML Kit on Android, Google (online), available from (https://developers.google.com/) (accessed 2021-09-02).

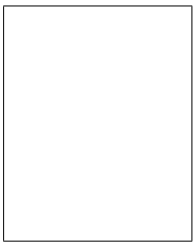
[14] Iowa Oral Performance Instrument(IOPI), IOPI Medical (online), available from (https://iopimedical.com/) (accessed 2021-09-02).

[15] Ito, K.: The Measurement of Oral Diadochokinesis with new measurement device. Niigata Dental Society Vol.39, No.1, pp.61–63 (2009).

[16] Kamijo, Y., Kanda, E., Ono, K., et al.: Low tongue pressure in peritoneal dialysis patients as a risk factor

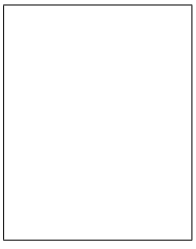
for malnutrition and sarcopenia: a cross-sectional study. Renal Replacement Therapy, Vol.4, pp.1-8 (2018).
[17] King, D.E.: Dlib-ml: A Machine Learning Toolkit. Journal of Machine Learning Research, Vol.10, pp.1755–1758 (2009).
[18] Mariakakis, A., Banks, M.A., Phillipi, L., et al.:BiliScreen: Smartphone-Based Scleral Jaundice Monitoring for Liver and Pancreatic Disorders. Proc. ACM Interact. Mob. Wearable Ubiquitous Technologies, Vol.1, No.2, pp.1-26 (2017).
[19] Mohamed, R., and Youssef, M.: HeartSense: Ubiquitous Accurate Multi-Modal Fusion-based Heart Rate Estimation Using Smartphones. Proc. ACM Interact. Mob. Wearable Ubiquitous Technologies, Vol.1, No.3, pp.1-18 (2017).
[20] Oguchi, K., Saitoh, E., Mizuno, M., et al.: The Repetitive Saliva Swallowing Test(RSST)as a Screening Test of Functional Dysphagia. The Japanese Journal of Rehabilitation Medicine, Vol.37, No.6, pp.375–382 (2000).
[21] Papapetropoulos, S. and Singer, C.: Eating dysfunction associated with oromandibular dystonia: clinical characteristics and treatment considerations, Head & Face Medicine, Vol.2, No.47, pp.1-4 (2006).
[22] GOKURI, 人工知能が嚥下を測る, PLIMES (オンライン), 入手先 (https://www.plimes.com/gokuri) (参照 2021-09-02).
[23] Rashid, H., Mendu, S., Daniel, K.E., et al.: Predicting Subjective Measures of Social Anxiety from Sparsely Collected Mobile Sensor Data. Proc. ACM Interact. Mob. Wearable Ubiquitous Technologies, Vol.4, No.3, pp.1-24 (2020).
[24] Sakai, K., Gilmour, S., Hoshino, E., et al.: A Machine Learning-Based Screening Test for Sarcopenic Dysphagia Using Image Recognition. Nutrients, Vol.13, No.11, p.4009 (2021).
[25] Shih, C.H., Tomita, N., Lukic, Y.X., et al.: Breeze: Smartphone-based Acoustic Real-time Detection of Breathing Phases for a Gamified Biofeedback Breathing Training. Proc. ACM Interact. Mob. Wearable Ubiquitous Technologies, Vol.3, No.4, pp.1-30 (2019).
[26] Tomasi, C. and Kanade, T.: Detection and Tracking of Point Features. International Journal of Computer Vision, Vol.9, pp.137-154 (1991).
[27] Dysphagia: Swallowing Disorders, U.S. National Library of Medicine NIH (online), available from ¡https://medlineplus.gov/swallowingdisorders.html¿ (accessed 2021-09-02).
[28] Wang, E.J, Li, W, Hawkins, D., et al.: HemaApp: noninvasive blood screening of hemoglobin using smartphone cameras. Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16), pp.593-604, ACM (2016).
[29] Wang, E.J., Zhu, J., Jain, M., et al.: Seismo: Blood Pressure Monitoring Using Built-in Smartphone Accelerometer and Camera. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems(CHI '18), pp.1-9, ACM (2018).
[30] Wang, R., Wang, W., Aung, M.S.H., et al.: Predicting Symptom Trajectories of Schizophrenia using Mobile Sensing. Proc. ACM Interact. Mob. Wearable Ubiquitous Technologies, Vol.1, No.3, pp.1-24 (2017).
[31] Zhang, H., Xu, C., Li, H., et al.: PDMove: Towards Passive Medication Adherence Monitoring of Parkinson's Disease Using Smartphone-based Gait Assessment. Proc. ACM Interact. Mob. Wearable Ubiquitous Technologies, Vol.3, No.3, pp.1-23 (2019).
[32] Zhu, M., Shi, D., Zheng, M., et al.: Robust Facial Landmark Detection via Occlusion-Adaptive Deep Networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp.3481–3491 (2019).
[33] 秋山理加, 濱嵜朋子, 酒井理恵, 片岡正太, 角田聡子, 邵仁浩, 巴美樹, 粟野秀慈, 岩崎正則, 安細敏弘. 介護施設利用高齢者における簡易嚥下状態評価票 (EAT-10) と口腔内環境, 口腔機能, 栄養状態との関連. 口腔衛生学会雑誌, 68(3): 128-136 (2018).
[34] 耿世嫻, 平井雄太, 下島銀士, 柳田陵介, 山田大志, 小野寺宏, 戸原玄, 矢谷浩司: スマートフォンを用いた画像認識による口腔・嚥下機能の定量的評価手法, マルチメディア, 分散, 協調とモバイルシンポジウム 2022 論文集, pp. 1141-1148 (2022).
[35] 平井雄太, 耿世嫻, 下島銀士, 小野寺宏, 矢谷浩司: 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, 研究報告ユビキタスコンピューティングシステム (UBI) 2021, No.38, pp.1-8 (2021).
[36] 西尾正輝: AMSD 標準ディサースリア検査 (Assessment of Motor Speech for Dysarthria), インテルナ出版 (2004).
[37] 西尾正輝, 阿部尚子, 岡本卓也, 福永真哉: 標準ディサースリア検査の嚥下障害への臨床的応用の試み:AMFD の開発, Japan Journal of Clinical Research in Dysarthria, Vol.6, No.1, pp.4-10 (2016).
[38] JMS 舌圧測定器 TPM-02, 株式会社ジーシー (オンライン): 入手先 (https://www.gcdental.co.jp/sys/data/item/1197/) (参照 2021-09-02).

付　　録



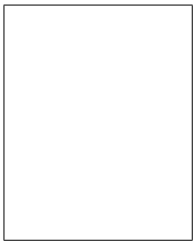
耿 世嫻

1997 年生．2022 年東京大学大学院工学系研究科電気系工学専攻修士課程修了．2022 年東京大学大学院工学系研究科電気系工学専攻博士課程進学．モバイルヘルス, デジタルメンタルヘルスの研究に従事する．



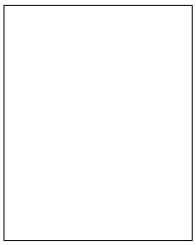
平井 雄太

1996 年生．2020 年東京大学工学部電気電子工学科卒業．2022 年東京大学大学院工学系研究科電気系工学専攻前期博士課程修了．2022 年 (株) 東芝入社，無線システムの研究に従事する．電子情報通信学会会員．



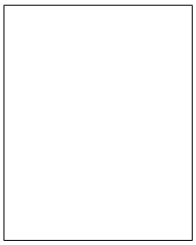
下島 銀士 （学生会員）

2000 年生．2022 年東京大学工学部電子情報工学科卒業．2022 年東京大学大学院学際情報学府学際情報学専攻修士課程入学．ヘルスケア・デジタルメンタルヘルスの研究に従事する．



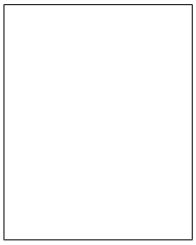
柳田 陵介

1995 年生．2019 年東京歯科大学歯学部卒業．2020 年東北大学病院歯科臨床研修プログラム修了．2020 年東京医科歯科大学大学院医歯学総合研究科博士課程入学．2022 年タイ・Naresuan 大学留学．2023 年東京医科歯科大学病院医員．摂食嚥下リハビリテーション学の診療および研究に従事．International Association for Dental Research，日本老年歯科医学会，日本摂食嚥下リハビリテーション学会各会員．



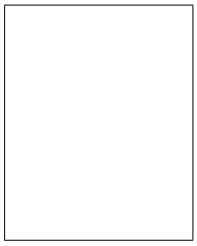
山田 大志

1992 年生．2019 年北海道大学歯学部卒業．2020 年東京医科歯科大学歯学部附属病院臨床研修プログラム II 修了．2020 年東京医科歯科大学大学院医歯学総合研究科摂食嚥下リハビリテーション学分野博士課程, 日本老年歯科医学会会員, 日本摂食嚥下リハビリテーション学会会員, 日本頭頸部癌学会会員



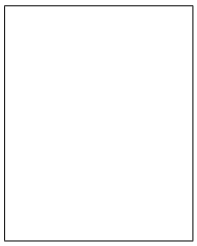
小野寺 宏

1956 年生．1981 年東北大学医学部卒業, 1985 年東北大学博士課程修了（医学博士），日本内科学会会員，日本神経学会会員



戸原 玄

1972 年生．1997 年東京医科歯科大学卒．2002 年同大学大学院博士課程修了, 歯学博士．2003 同大学医員．2005 年同大学助手．2008 年日本大学准教授．2013 年東京医科歯科大学高齢者歯科学講座准教授，2020 年同大学摂食嚥下リハビリテーション学分野教授．



矢谷 浩司 （正会員）

2014 年より東京大学大学院工学系研究科電気系工学専攻准教授．同大学にてインタラクティブ・インテリジェント・システム ラボ (https://iis-lab.org) を率いる．デジタルヘルスケア，ユーザブルセキュリティ，生産性・創造性支援，インタラクティブシステムにおける AI 技術の創造的利用や新しいセンシング技術の研究に従事．ACM Distinguished Member．