

嚥下機能検査のための嚥下区間自動検出システムの開発と評価

鈴木 翔大[†] 古川 大輔[‡] 黒岩 眞吾^{††} 西田 昌史[†] 西村 雅史^{†††}静岡大学[†] 君津中央病院[‡] 千葉大学^{††} 愛知産業大学^{†††}

1. はじめに

加齢に伴う口腔機能の低下は誤嚥性肺炎のリスクを高めており、嚥下機能低下の早期発見は、健康寿命を延伸する上で重要な課題である。現在の臨床現場では言語聴覚士が摂食場面に立ち会い、患者の嚥下音を直接聴取することで嚥下の質を評価しているが、この方法はリアルタイムでの評価に限られ、後日の客観的な検証や多職種間での情報共有が困難である。我々は、録音データから嚥下を自動検出し、言語聴覚士が事後的に嚥下音を効率的に評価できる支援システムの開発を進めている。

本研究では、実際の検査環境データの事後的な評価作業における提案システムの有効性を検証するとともに、効率の改善の鍵となる嚥下自動検出性能の改善方法について検討したので報告する。

2. 提案システム

2.1. 嚥下区間確認システム

本研究で対象とした病棟の食事場面では、言語聴覚士が皮膚接触型マイクで収録された患者の嚥下音を聴取するとともに、その食事音を同時に録音している。今回はその録音データから嚥下区間の自動検出を行い、検出した嚥下区間を言語聴覚士が確認、修正可能なシステムを構築した。必要とされる機能については言語聴覚士のアドバイスに基づいて作成した。

動作画面を図1に示す。入力となる食事音ファイルを選択し、嚥下検出を実行することで音声の波形図と嚥下区間が可視化される。システム上で食事音の再生や、嚥下区間の追加・削除が可能であり、図1下部から嚥下区間編集画面に遷移して、嚥下区間の位置修正や、嚥下音の質を主観評価する機能も実装した。システムを用いた処理の手順を図2に示す。まず、皮膚接触型マイクによって録音された食事音を用意し、それをシステムに入力して嚥下音を自動検出し、結果の表示までを行う。その後、検出された各嚥下区間に対して、実際に嚥下音を聞きながら必要に応じて区間の修正を行い、その後、さらにその嚥下音を聴取して、診断を行う流れを想定している。診断後は、嚥下区間のみを切り出して結果を保存することが可能であり、診断結果の事後的な確認にも有効である。

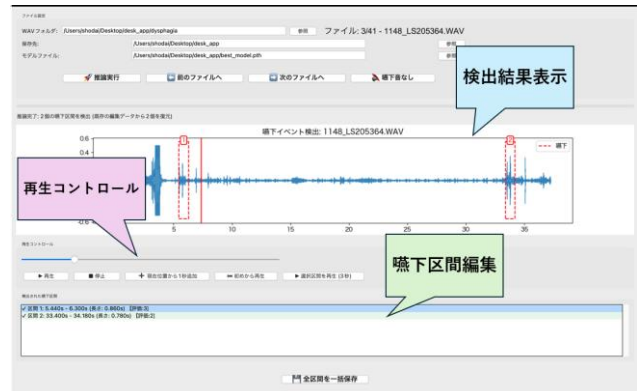


図1 嚥下区間確認システムの動作画面

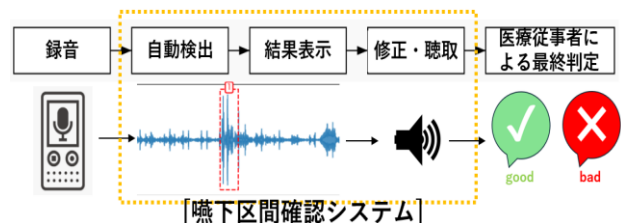


図2 システムを用いた処理の動作手順

2.2. 食行動認識モデル

これまでに非侵襲な嚥下自動検出法として、生体音を活用した食行動認識モデルが提案されている[1][2]。しかし、これらの研究は防音室等の管理環境下での評価や、嚥下障害のない健常者の嚥下音での評価にとどまっており、実際の検査環境との乖離が存在する。嚥下区間認識システムのベースモデルとして、自己教師あり学習モデルを用いた嚥下区間自動検出モデルを利用した[2]。特徴抽出器を WavLM とし、3 クラス {咀嚼, 嚥下, 発話} の時系列分類に GRU を利用したモデルに、食行動音と Commonvoice 日本語音声を学習させたモデル (WavLM+GRU_CV) である。実用的なシステム構築のためには、ベースモデルの嚥下区間検出性能の向上が必要であった。そこで、本研究では病院や実環境データを用いて Fine tuning を実施し、更なる精度改善を図った。具体的には、嚥下障害者の食事音と、健常者の日常食事音を用いて 2 段階の Fine tuning を実施したものを提案システム用のモデル (Fine-tuned_Model) とした。

3. 評価実験

3.1. Fine tuning 用データ

ベースモデルの Fine tuning 用データとしては、君津中央病院で収録された嚥下障害者と、静岡大学で収録された若年健常者の複数人での会話を含む食

Development and Evaluation of an Automatic Swallowing Segment Detection System for Swallowing Function Assessment

[†]Shodai Suzuki [‡]Daisuke Furukawa [‡]Shingo Kuroiwa

[†]Masafumi Nishida ^{††}Masafumi Nishimura

[†]Shizuoka University [‡]Kimitsu Chuo Hospital

^{††}Chiba University ^{†††}Aichi Sangyo University

行動時の収録音データを用いた。具体的には、病院データには、嚥下音 128 回を含む 30 名の嚥下障害者のゼリーの摂食場面の約 40 分を利用し、大学データには、嚥下音 537 回を含む 7 名の健常者のピザ、リンゴ、水を対象とした摂食場面の約 150 分を利用した。

3.2. 評価データ

モデルの評価には、病院で収集された入院患者の病棟でのペースト食の摂食場面の収録音を利用した。具体的には、病院の収録音データのうち、学習データに含まれていない、嚥下機能の異なる 3 名(健常者 2 名、嚥下障害者 1 名)の食事音データである。このデータには入院患者 3 名の食事の開始から終了までの食事音延べ約 40 分間が収録されている。ここでは、入院患者のうち、他の疾患はあるが嚥下機能は正常な患者を健常者、嚥下機能に障害を持つ患者を嚥下障害者としている。評価データには、食事の合間に発話や咳といった食事音以外の音も含まれており、嚥下音は計 340 回存在する。

3.3. 評価指標

嚥下の検出精度を Precision, Recall, F1 スコアにより評価した。嚥下検出の評価には、予測区間と正解区間の重複度を示す IOU (Intersection over Union) しきい値 (0.01) を用い、正解区間と予測区間が一部でも重複すれば正解とした。

表 1 健常者での嚥下検出精度

モデル	Precision	Recall	F1-score
WavLM+GRU_CV	0.725	0.983	0.833
Fine-tuned_Model	0.973	0.967	0.970

表 2 嚥下障害者での嚥下検出精度

モデル	Precision	Recall	F1-score
WavLM+GRU_CV	0.858	0.885	0.872
Fine-tuned_Model	0.922	0.947	0.935

3.4. 実験結果

表 1 に 2 名の健常者の評価データの嚥下区間検出精度を、表 2 に嚥下障害者の食行動音に対しての嚥下区間検出精度を Precision, Recall, F1 スコアで示す。健常者と嚥下障害者のいずれのデータに対しても、今回構築した Fine-tuned_Model がベースモデルよりも高い検出性能を示した。具体的には、健常者データにおいて適合率が、嚥下障害者データにおいて適合率と再現率が向上した。Fine tuning によりノイズなどによる嚥下の誤検出や、嚥下障害者の嚥下音の見逃しが抑制された。結果から、構築したモデルは病床での収録環境や、被験者の嚥下機能に関わらず、高い精度で嚥下区間を検出できることが示唆された。

図 2 に、評価データの食事音に含まれる音響イベントの時間的割合を示す。なお、その他の区間には無音や分類不可能な雑音などが含まれている。内訳を分析した結果、最も多く含まれていたイベントは

発話であった。入院患者、特に嚥下障害者は食べ物を誤嚥する可能性から、摂食介助者が患者に対して正しく食事できているかを確認する必要があるため、食事時の収録音でも発話が多く含まれている。他には、嚥下音が収録音の約 9% を占めており、咳が 4% 含まれていた。

このように、病棟で収録された食事音データには、調査対象である嚥下音は時間的割合で 1 割程度しか含まれておらず、発話やその他音声が大半を占めている。今回提案したシステムを用いることで、嚥下音の確認から評価までを効率的に実施できると考えられる。

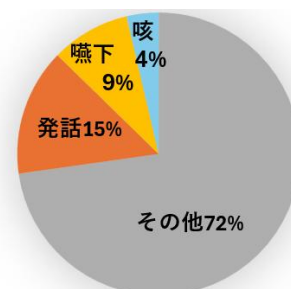


図 3 評価データ内の音響イベントの内訳

4. おわりに

本研究では、長時間の食事音データから嚥下区間候補を自動検出することで言語聴覚士が、事後的に嚥下音区間の修正及び主観評価を効率よく実施できるシステムの開発を行なった。また、実際の検査環境下データを用いた Fine tuning によって高い検出性能が得られることを確認した。今後は嚥下音の音響特徴から嚥下機能の評価するシステム[3]との統合を進め、最終的には食事音から嚥下の認識、嚥下機能の評価までを自動で行い、誤嚥リスクを予測するシステムの開発を目指す。これにより、専門家でなくても一定水準の嚥下評価が可能となり、誤嚥性肺炎の予防と医療者の負担軽減に貢献できると考えている。

謝辞

本研究は静岡大学及び君津中央病院での倫理審査を受けて承認されました(承認番号 P0001729988)。また、本研究の一部は JSPS 科研費 21K18305 の助成を受けています。

参考文献

- [1] Y. Khalifa, C. Donohue, J. L. Coyle, and E. Sejdic, "Autonomous swallow segment extraction using deep learning in neck-sensor vibratory signals from patients with dysphagia," IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 2, pp. 956-967, 2023.
- [2] Toshihiro Tsukagoshi, Masafumi Nishida, Masafumi Nishimura, "Swallowing Sound Segmentation Using Self-Supervised Learning-Based Features" Proc. of GCCE, pp.770-772, 2025.9.
- [3] Ryo Ushijima, Daisuke Furukawa, Shingo Kuroiwa, Masafumi Nishida1, Masafumi Nishimura "Visualization and Classification of Swallowing Difficulties Using Scalogram-Based CNN" proc. of RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing, pp.356-359, 2025.