

高齢者の食べることと話すことを支える情報システムに関する研究

代表研究者	西村 雅史	静岡大学 学術院情報学領域 教授
共同研究者	桐山 伸也	静岡大学 学術院情報学領域 准教授
共同研究者	西田 昌史	静岡大学 学術院融合グローバル領域 准教授

1 はじめに

超高齢社会が現実となり、高齢者の心身の状態、さらにはその家族や生活環境までを含めた関わりを幅広くかつ正確に把握することで、高齢者の健康管理と介護サービスの効率化を同時に実現したいという要望がある。その中でも特に高齢者の「食べること」は嚥下の能力に、また、「話すこと」は認知能力に関わる重要な行動要素である。「食べること」の能力低下は高齢者の身体機能低下の大きな要因となっており、低栄養状態に起因する筋肉の消耗、そして、さらに重度な嚥下障害といった負のスパイラルに陥ることが知られている。一方、「話すこと」は人間の知的活動に深く関係しており、会話量や人との関わりの減少は認知症発症のリスクを高めることが知られている。本研究では、このような「食べること」「話すこと」を中心とした高齢者の行動を非侵襲かつ長時間モニタリングするための手段として、1) 咽喉マイクと集音マイクを用いて2チャンネルの音データを収録する方法、2) 多人数から同時にそれらの音データをオンライン収集するシステムを開発した。また、3) 収集された音データを用いてフレーム単位もしくはイベント単位に局所的な行動識別を行う方法、4) それらの行動識別情報に基づいて「会話行動の詳細」を分析する方法、5) 「食事」「会話」といった大局的な行動を検出するための方法、をそれぞれ検討した。そして、6) 実環境を模擬した環境において高齢者グループの行動データを収集し、それぞれの方法の基本性能を評価した。結果として、会話行動についてはF値0.88の性能で、各人の発話区間を安定的に、かつ正確に検出できることが分かった。一方、「発話」、「咀嚼」、「嚥下」などの局所的な行動認識結果を統合することで、大局的な「食事」、「会話」といった行動の検出についても、安定した検出が可能となること、さらには「会話しながらの食事」といった複雑な行動についても検出・分析が可能となる見通しを得た。

2 行動音データの収録方法

2-1 行動音データ収集の問題点

ライフログの研究などではセンサー情報の一つとして周囲の環境音や音声を収録することが多いが、その情報が十分に活用されてきたとは言いがたい。特に通常のマイクでは対象者に関して音声以外の生体音情報を得ることは出来ないし、その音声も多くの場合口元から遠く離れた位置で周囲の人の発話や環境音と共に収録されることになる。結果としてノイズや反射等の影響を強く受けるため、音声認識技術などの適用は極めて困難であり、十分な情報を収集することが出来なかった。

2-2 咽喉マイクによる収録

本研究では上記のような問題点を考慮し、周囲ノイズの影響を受けにくいマイク構成として、咽喉マイクと通常集音用ピンマイクの併用を検討した(図1)。咽喉マイクは高騒音下の会話用に開発された接触型のマイクで、咽喉付近の皮膚の振動を音として収集するものである。このため収録音の帯域はおおよそ4kHz以下となり、こもった音に感じられるが、発話内容はおおむね問題なく理解できる。また、図1から分かるようにネックバンド型の形状をしており、装着は容易で、外れにくい。装着者の負担も比較的小さいので、就寝時を含む高齢者の長時間



図1 咽喉マイクと集音マイクの同時装着の様子

の装着も可能である。また、唾や飲食物を飲み込む際に咽喉付近で発生する嚙下音は高齢者の身体状態を知る上で重要な情報であることが知られており、医療分野でも咽喉マイクによる計測が行われている。このように、咽喉マイクは周囲ノイズの影響を受けにくい音声データの収集に加えて、嚙下音を中心とした生体音を感度良く収集出来るという特徴を持つ。同様の目的に使用されるマイクとしては骨伝導マイクが知られているが、通常、イヤホンかヘッドセットの形状をしており、意図せず外れてしまうことも多い上、耳穴をふさいでしまうこともあって、長時間の装着が困難な場合がある。また、嚙下音の観測が困難であるため、我々の目的には必ずしも適していない。なお、骨伝導と称されているマイクも、実態としては咽喉マイクと同じ肉伝道音の集音デバイスに分類される。

2-3 マイクによる同時収録の意義

集音マイクと咽喉マイクの両者を用いることで、嚙下などの生体音が収集出来ることに加え、特に音声については以下の効果が期待される。1) 本人を交えた会話か、他人の一方的な発話か、本人の独り言かの識別精度の向上。2) 咽喉マイクの信号を用いた集音マイク側の信号からの雑音除去[1]。特に1)は高齢者の知的活動を検出するために大変重要である。つまり、従来の集音マイクのデータでは音声を検出できたとしても、それが認知症の特徴でもある“独語”なのか、テレビ視聴などの状況なのかを識別することは不可能、もしくは大変困難であった。一方、咽喉マイクを併用することで、それぞれのマイクで検出された音声区間の重なり情報によって、通常の会話状態と、それ以外の状況を従来よりも簡単に、かつ正確に区別出来る可能性がある。

2-4 オンライン・データ収録システムの構成

図1に示す集音デバイスを行動分析の対象者全員に装着してもらい、スマートフォン上に実装した音声収録ソフトを用いて録音を実施する。なお、咽喉マイクには NANZU SH-12K を、集音マイクには SONY ECM CS3 を使用した。録音データを定期的にサーバーにFTP転送することでオンラインでの多人数データ収録も可能なシステムとしている。なお、多人数データ収録時のデバイス毎の時間同期については、FTP転送時のシステムクロックを付記することで事後的な調整を可能にしている。

3 学習用・評価用データベースの構築

3-1 収集行動音データ

3-1-1 実験室内行動音データ

20代男性3人組による約30分間の「会話音声」及び「摂食行動音」を、簡易防音室内で先の2チャンネル集音デバイス及びデータ収集システムを用いて収録した。この時、飲料及び菓子などを提供した上で、一定量の発話や飲食を所定時間内に必ず行うように指示した。このような収録を、対象者を変えて計4度実施し、のべ12名、計約2時間分の行動音を収録した。

3-1-2 実環境行動音データ

互いに顔見知りの65歳以上の高齢健常者6名が、エアコンの動作音、ドアの開閉音、テレビ音声などの騒音もある通常の家環境を模擬した実験室に集まり、先の2チャンネル集音デバイスを装着して行動音の収録を実施した。弁当の他に菓子や茶などの飲料も提供し、会話に加えて、食事や部屋の出入りなどの行動も自由に行ってもらった。収集は毎回午前10時から午後2時まで計4時間程度行い、これを複数の日に分けて実施し、のべ約120時間分の音を収録した。なお、各実験に参集した被験者数は最小で4名、最大6名であった。

3-2 データベースの構築

データ収集システムで収録されたデータをサーバー上に集約した。実験室内行動音データ2時間分については、「嚙下」「咀嚼」「発話」「雑音」の4種類のラベルを手で詳細に付与した。これらは検出された音区間(概ね25msec~1秒程度)ごとに付与されている。

評価用の実環境行動音について、1) 会話を中心とした1時間程度の区間と、2) 会話と食事が含まれた1時間程度の区間、それぞれに対し、同様に詳細な音区間ラベルの付与を行った。前者は主に会話状況下において特定話者の「話す」行為を正確に検出できるかを評価する目的で利用し、後者は主に「嚙下」「咀嚼」

「発話」といった局所的な行動の識別性能を評価する目的で利用する。

一方、「発話」「食事」という行動はこれらの局所的な行動の連鎖として観測されるため、別途一連の発話を「会話」、食事の始まりから終了を「食事」区間として大まかな行動をラベル付けする。今回は食事や会話を多く含むのべ8時間程度の評価用実環境行動音に対し、「会話」及び「食事」のラベルを手手で付与した。これは、「話す」と比較して「食べる」行為に関しては一連の摂食行動の有無の検出のみが重要であり、「咀嚼」「嚥下」といった個別の詳細行動の認識は必ずしも必要とされないと考えたためである。なお、ラベル付与の際には収録音データだけでなく、同時に収録した画像データも参考にしている。結果として、8時間分のデータのうち、約22%が会話区間、約21%が食事区間とラベル付けされた。

4 「話すこと」の自動検出

4-1 発話行動及び会話の検出

高齢者の知的活動を正確に把握する手段として、収録音声データから対象話者の発話及び、周囲話者の発話を正確に分離し、多人数の会話の様子を可視化する方法について検討を行った。ここでは、単に「話している」という動作を検出するだけでなく、「どう話しているのか」という、会話行動の詳細を明らかにすることを目標とする。例えば、高齢者が会話に参加していたとしても、単に会話の場にいただけなのか、内容を理解して相槌などを打つ程度の会話だったのか、あるいは、積極的に発言を行った会話だったのかという区別は、高齢者の知的活動を知るうえで大変重要と考えられるからである。この目的を達成するためには、数百ミリ秒の継続時間しかない相槌も、正確に検出を行う必要がある。

4-2 提案手法

通常のマイクで収集される音には、対象者の発話以外に、会話に参加している人たちの発話や、周囲の雑音・騒音などが含まれ、特に高音音下や発話衝突（発話者の音声の被り区間）が生じる場合には対象者の発話区間を正確に検出するのは困難とされている。一方、咽喉マイクを用いれば、周囲雑音の影響をほとんど受けることなく対象話者の発話を検出できることがわかっているが、実際には衣摺れ音や嚥下音などの生体音が雑音となって誤検出を生じるため、特に相槌などの短時間の発話を正確に区間検出することは難しい。ここでは咽喉マイク側の入力を GMM (Gaussian Mixture Model) によるスペクトル特徴量のモデル化によって識別し、これらの雑音、特に衣摺れ音の影響を効率的に排除するとともに、ピンマイク側から得られる音区間情報も併用することで、高精度で、発話衝突にも頑強な発話検出を試みる。これによって、特に、会話において重要な役割を果たす相槌などの持続時間の短い発話に対しても高精度の検出が可能になる。

4-2-1 咽喉マイク収録音に対する音区間検出

学習用の音データに対して、発話区間には speech、それ以外の特に嚥下や咳等のイベントが起こっていない区間を非発話区間として sil のラベルを手手で付与し、当該区間のデータを用いてそれぞれの GMM を学習した（いずれも混合分布数 32）。特徴量には、窓幅 25 msec、フレーム周期 10 msec で抽出した MFCC (Mel-Frequency Cepstrum Coefficient) を用い、0 番目のケプストラム係数を含めた低次から 13 次元、その Δ 、 $\Delta\Delta$ の計 39 次元の特徴量を使用した。なお、嚥下や衣摺れといった雑音を事前学習することでより正確に発話区間の検出が可能になると考えられるが、今回は学習データ量が限られていたこともあり、むしろ性能低下が見られたので雑音は学習には含めないこととした。

一方、評価時には、フレームごとに各 GMM の尤度を比較することで非発話区間か発話区間であるかの判定を行う。

4-2-2 ピンマイク収録音に対する音区間検出

ピンマイクによって収録される環境音に対してはパワー情報による音区間検出を行う。具体的には窓幅 40 ms、フレーム周期 20 ms とし、フレーム毎の対数パワー情報に対して、閾値処理を行った。なお、入力データに対し、事前に対数パワー値のヒストグラムを取り、非音声区間のモード値と音声区間の 90 パーセント値の間を 100 とし、入力対数パワー値の正規化を実施している。この操作によって固定閾値でも安定した音区間検出が可能となっている。

ここで GMM のような統計モデルを用いず、パワーベースの音区間検出としたのは、集音マイクで収集される音が非常に多様であって、限られた学習データでモデルを構築しても、十分な識別性能が得られないと考えたためである。一方、パワー値に基づいて音の有無を判定し、この情報を咽喉マイク側の検出・識別結果

と統合することで、より高精度な検出を行う。

4-2-3 推定音区間のスムージング

GMM 及びパワー情報による音区間検出では、フレーム毎に逐一発話区間と非発話区間の判定を行うため、ごく短い区間で発話区間と非発話区間が交互に検出されることが多い。これを防ぐため、推定発話区間に対してスムージングを行う。微小時間の判定揺れや嚙下などの雑音を発話とした区間を消去し、かつ実際の発話区間に影響がないよう、暫定的に発話区間に関する閾値を 0.2[sec]、非発話区間に関する閾値を 0.3[sec] と定め、閾値以下の区間を削除し、前後区間の結合を行った。この処理は各マイクに対する発話区間推定結果それぞれに行い、また推定結果を統合した後にも再度行っている。

4-2-4 2チャンネル情報の統合

各マイクで推定された発話区間に対し、統合処理を行う。この際、両方の音から発話区間と推定された区間のみを推定発話区間とした(図2)。

咽喉マイクの音に対して GMM で推定した区間が、ピンマイクのパワー情報により推定した区間に無い場合はその区間を削除し(図2, A)、ピンマイクの音から他人の発話が発話区間と推定されても、咽喉マイクで発話区間として検出されていなければ削除できる(図2, B)。この手法により、ピンマイクで発話区間と推定される他人の声や外部雑音、咽喉マイクで発話区間と推定される雑音の両方を推定区間から削除することができる。さらに、両マイクでの誤推定区間が重複してしまった場合でも、重複区間が短ければ統合処理において取り除くことができる(図2, C)。

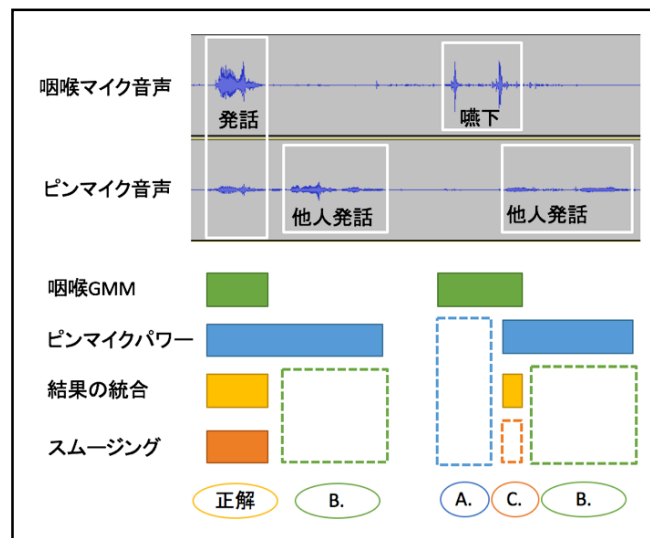


図2 2チャンネル情報の統合による発話区間推定

4-3 実験室内行動音データによる評価

20代男性3名による5分間の自由発話を1セッションとし、合計5セッションを行って、計887発話を収集し、基本性能を評価した。このデータは3-1-1に記した実験室内行動音とは別データの、会話のみのデータである。一方、GMMの学習にはこの3名とは異なる3名の話者による会話音声(計138発話)を用いている。ピンマイクの信号に対してパワーベースの音区間検出を行なった結果、咽喉マイク側の信号に対してGMMによる識別を適用した場合の結果、並びに、両方の検出結果を統合した提案手法の結果を図3に示す。なお、音区間検出性能の評価はフレーム単位ではなく、音イベントごとの検出・識別性能である。正解区間と比べて推定音区間が0.5秒以上ずれていた場合は不正解とし、再現率、適合率、及びF値を用いて評価した。この結果から、ピンマイク側の検出結果の情報も統合利用することで、特に適合率を大きく改善できることが分かった。なお、咽喉マイクだけを使った場合と提案手法の比較の形で、検出誤り原因の内訳を図4に示す。ここに示すように、咽喉マイクとGMMだけでは発話区間に誤認識された衣摺れ音などのケースについて、提案方法が有効に動作したことが分かる。

4-4 実環境行動音データによる評価

3-1-2 に示した実環境行動音データのうち、5名の高齢者による1時間の自由会話部分を抽出し、10分をGMMの学習、残りの50分(2,524発声)を評価データとして、提案手法の検証を行なった。結果としてF値=0.88が得られ、実環境の行動音データに対しても提案手法が有効に動作することを確認し、会話の参加状況や会話主導権の可視化の見通しを得ることができた。

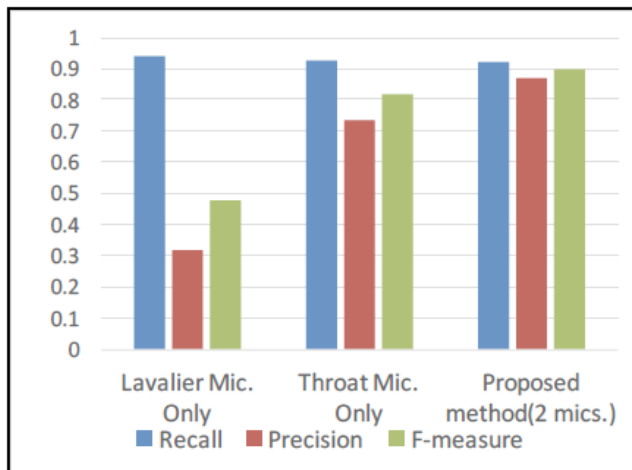


図3 咽喉マイク単体と提案手法の発話区間推定性能比較

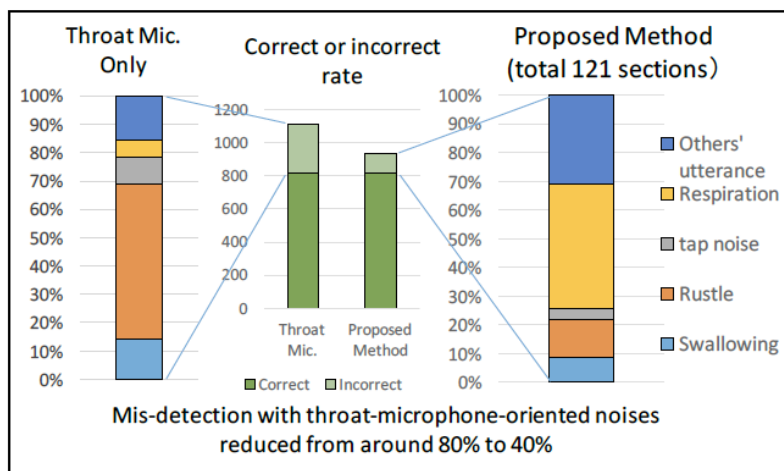


図4 音声区間検出誤り原因の比較

5 「話すこと」「食べること」に関する行動自動検出

5-1 大局的行動の検出

これまでは「話すこと」に焦点を当て、会話状態の詳細な分析が可能となるような、マイクロな単位での発話行動の検出を行なってきた。一方、ここでは発話内容の詳細な分析を目指すのではなく、「十分な量の会話をしているか」、「ちゃんと食事をしているか」といったレベルで、高齢者の行動を検出する方法についての検討を行う。特に、個別の行動音、特に食事に関連する音は多様であり、自動化のためのモデル学習には相応の量の学習データが必要になる。実際に、DNNを用いた行動認識の実験を行ったところ、今回用意した程度のデータ量では過学習となる傾向が見られ、十分な性能が得られなかった。また、「食事」という行動区間

自体は、「咀嚼」「嚥下」あるいは一部に「会話」や「雑音」も含めた区間として認識される。このため、フレーム単位(10msec)でGMMによる「発話」「咀嚼」「嚥下」「雑音」の識別を行ったのち、これらの識別結果を統合処理することが必要になる。また、「会話」区間についても、フレーム単位の各チャンネルの識別結果を統合することでさらなる高精度な検出を実現できる可能性もある。

5-2 提案手法

ここでも「話すこと」のみの自動検出時と同じく2チャンネルの集音デバイスと多人数の同時収録システムを用いる。

これまで「歩行」「睡眠」といった大局的行動の認識においては、1分程度の固定時間内で最も高頻度に出現するフレーム単位の行動認識結果をその区間の行動とするといったルールが用いられることが多かったが、「食事」のような複雑な詳細行動の組み合わせで構成される大局的行動をこのような単純なルールで検出することは難しい。また、「話すこと」に限定したケースでは、咽喉マイクとピンマイクの2種類の情報をヒューリスティックルールに基づいて統合することでその精度を改善できることを示したが、多元的な情報に対して、このようなルールを構築するのは容易ではない。さらには、「食べること」は単独で起きケースだけではなく、「話すこと」や他の大局的な行動と同時に起きる可能性も高い。そこで、本研究ではLSTM(Long-Short Term Memory)と呼ばれるRNN(Recurrent Neural Network)を活用して、そのようなヒューリスティックルールに代わる結果統合処理を自動学習させる方法を提案する。なお、GMMによるフレーム単位の学習には詳細な行動ラベルが付与された学習データが必要であるが、この行動統合用のLSTMの学習にはフレーム単位の自動認識結果と、最終目的である「食事」「会話」等の大まかなラベルが付与されていれば良く、学習も効率よく行える。

5-2-1 咽喉マイク収録音とピンマイク収録音を用いたフレーム単位の行動識別

ここでは2チャンネルの入力信号それぞれに対して、窓幅25msec、フレーム周期10msecで分析し、4-2-1の実験と同様に、39次元からなる特徴量(MFCC+ Δ + $\Delta\Delta$)を推定した。ピンマイクに対しては「発話」「雑音」「無音」の3つの状態を、咽喉マイクに対してはこれに加えて「咀嚼」「嚥下」の5つの状態をそれぞれGMMで表現し、フレーム毎に最大尤度を返すGMMに対応する状態を局所的な行動識別結果とする。

5-2-2 LSTMによる局所行動識別結果の統合

前段でピンマイク及び咽喉マイクのそれぞれの入力に対してGMMによるフレーム単位の局所的行動識別を行なった結果をLSTMに入力し、「食事」「会話」といった複合的かつ大局的な行動の識別を行う。ただし、ここでは局所行動識別結果における「無音」及び「雑音」は同じものとして扱った。

LSTMを用いたのは、過去の識別結果を十分考慮して統合処理を行う必要があると考えたためである。LSTMは過去の情報を考慮するRNNの拡張として提案されたモデルで、RNNの中間層のユニットを、過去の状態を記憶するメモリセル及びその記憶期間を制御するゲートからなるLSTMブロックに置き換えることで実現されており、RNNの課題であった勾配消失・勾配爆発現象を抑えることができる。本研究ではForget Gateを導入したLSTMを利用した。中間層は1層でLSTM層となっている。出力層は3つのノードを持ち、それぞれ、「食事」「会話」並びに「それ以外の状態」に対応している。ここでは3つのノードのうち、確率値が最大となったノードに対応する行動を最終的な出力とした。この処理の流れを図5に示す。

なお、LSTMの学習には3-2に記した8時間分の学習ラベルと、そのデータに対する局所行動の自動識別結果を用いた。LSTM自体はフレーム毎に行動を出力するが、大局的な行動ラベルを正解データとして学習することによって、概ね大局的な行動を検出できるものと考えた。

5-3 実環境行動音データを用いた局所行動の識別性能評価

3-2で構築した1時間分の実環境行動音評価データを用い、咽喉マイク、ピンマイクそれぞれに対して、局所行動の識別性能を評価した。なお、GMMの学習には同様に、3-2に記した2時間分の実験室内行動音データを用いた。GMMの混合数は256とし、共分散行列は対角成分のみとしている。

なお、評価実験結果を図6、図7に示す。「発話」に関してはピンマイクも、咽喉マイクもフレーム単位の識別性能に大きな差はなく、F値=0.8弱程度である。また、咽喉マイクによる「嚥下」識別性能はF値=0.39、「咀嚼」では0.52と、これだけでは十分な識別性能が得られないこともわかった。なお、この評価結果はフレーム単位の評価であり、「話すこと」の検出において行なった発話区間単位の評価ではないことに注意されたい。

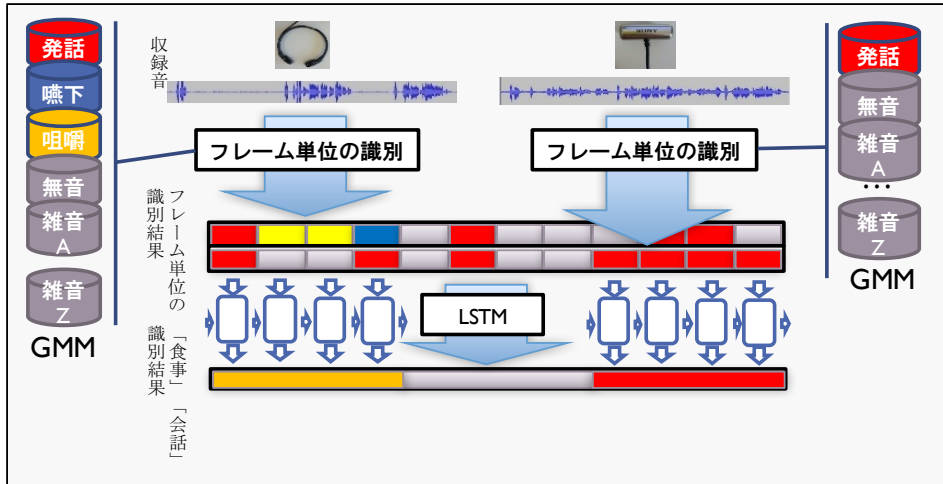


図5 LSTMによる局所行動識別結果の統合処理フロー

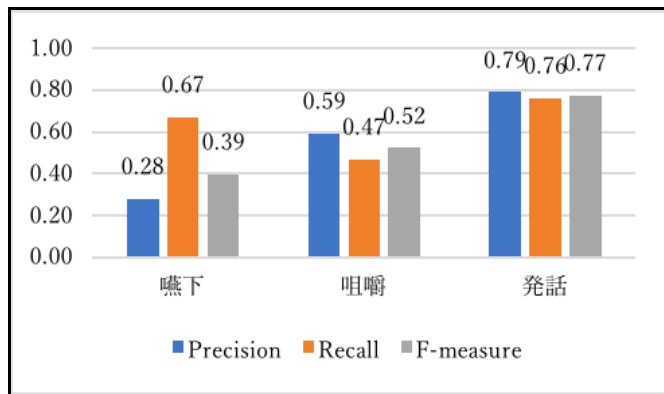


図6 咽喉マイクによる局所行動識別性能

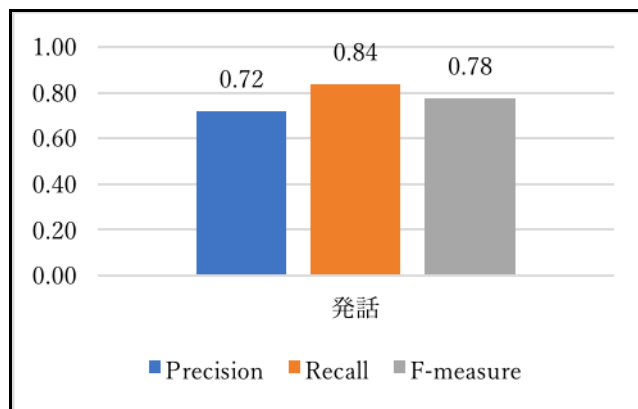


図7 ピンマイクによる局所行動識別性能

5-4 実環境行動音データを用いた大局的行動の識別性能評価

局所行動識別評価と同じ1時間分の評価データに対して、LSTMを用いた統合処理を行なった場合の結果を図8に示す。なお、比較のため、ヒューリスティック・ルールに基づいて局所行動識別結果を統合した場合

の結果についても示す。局所的行動の識別結果は不十分であったが、食事行動をF値=0.55で、会話行動もF値=0.77で識別できることがわかった。なお、今回評価に用いた1時間分のデータのうち、30分程度について、大局的行動の正解ラベル、LSTMによる識別結果、ヒューリスティック・ルールに基づく識別結果を図9に示す。特に、「食事しながら会話」といった、複雑な行動が観察された部分について、かなり人間の感覚に近い形で「食事」や「会話」といった行動を検出することができたと考えている。

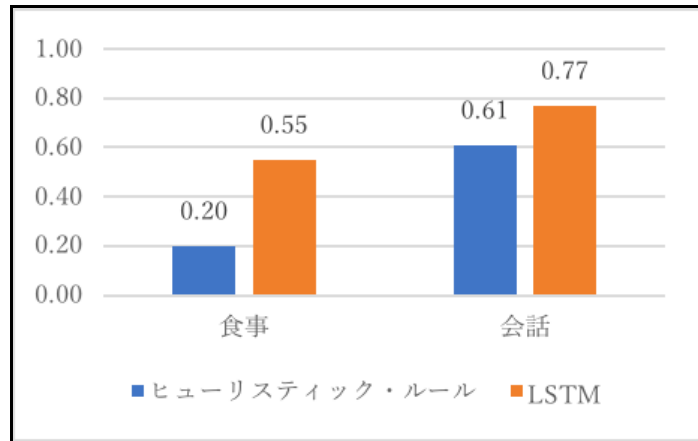


図8 「食事」「会話」の識別性能評価

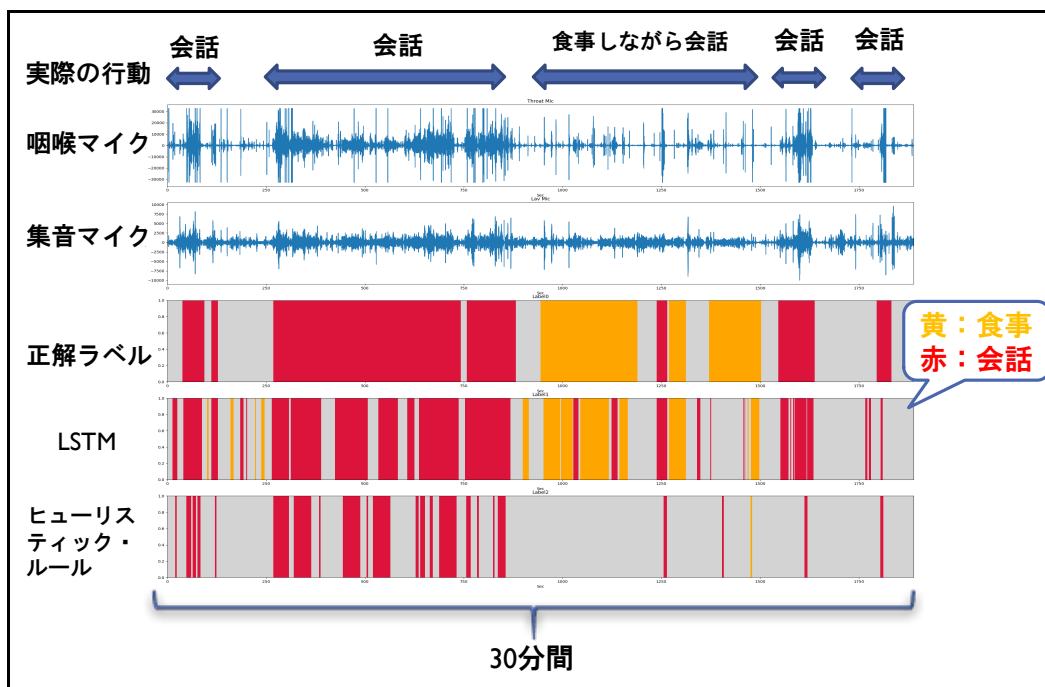


図9 「食事」「会話」の識別結果（30分間の実環境行動データ）

6 おわりに

高齢者の行動、中でも特に高齢者の「食べること」と「話すこと」に関連する行動を認識する目的で、音に着目し、非侵襲かつ簡便な集音デバイス及び収録システムを開発した。この音収集システムを活用し、「話すこと」に関する行動を自動検出するための方法を提案し、「相槌」を含む詳細な発話行動の検出を可能にした。この情報を用いることで、単に会話の場にいるだけなのか、内容を理解して相槌等を打っているの

か、さらには積極的に発言を行っているのかという詳細な会話状況の分析につなげることが出来る。今後は相槌だけでなく、「笑い」や「ため息」など、場の雰囲気や決定する要因についても自動抽出することで、さらに深い会話状況理解を実現したい。

一方、「食べること」「話すこと」に関する大局的な行動情報についても、2チャンネルマイクの情報をLSTMで統合すれば、局所的な行動の識別結果に基づいて、より正確な行動推定が可能になることを示した。特に、「食べながら話す」といった行動の識別可能性を示すことができた。今後はこの方法を改良して、「テレビを見ながら食事をしている」といった、より複雑な行動へと対象を広げたい。

【参考文献】

- [1] 玉森聡 他, "Recurrent Neural Network に基づく日常生活行動認識." 電子情報通信学会技術報告, Vol.116, No.189, pp.7-12 (2016)
- [2] 大内一成 他, "携帯電話搭載センサによるリアルタイム生活行動認識." 情報処理学会論文誌, Vol.53, No.7, pp.1675-1686 (2012)
- [3] Zhang, Zhengyou et.al., "Multi-sensory microphones for robust speech detection, enhancement and recognition." ICASSP2004, pp.781-784 (2004)
- [4] Bi, Yin et.al., "Autodietary: A wearable acoustic sensor system for food intake recognition in daily life." IEEE Sensors Journal, Vol.16, No.3, pp.806-816 (2016)
- [5] Ying, Dongwen et.al., "Voice activity detection based on an unsupervised learning framework." IEEE Transactions on Audio, Speech, and Language Processing, Vol.19, No.8, pp.2624-2633 (2011)
- [6] Gers, Felix A et.al., "Learning to forget: Continual prediction with LSTM." Neural computation, Vol.12, No.10, pp.2451-1471 (2000)
- [7] Sutskever, Ilya et.al., "Sequence to sequence learning with neural networks." Advances in neural information processing systems., pp.1-9 (2014)
- [8] X. Anguera, et.al., "Speaker Diarization: A Review of Recent Research" In IEEE Transactions on Audio, Speech, and Language Processing, vol. 20, issue 2, pp. 356-370. (2012)
- [9] 河原達也 他, "音声会話コンテンツにおける聴衆の反応に基づく音響イベントとホットスポットの検出", 情報処理学会論文誌, Vol.52, No.12, 3363-3373, (2011)

〈発表資料〉

題 名	掲載誌・学会名等	発表年月
咽喉音に基づく身体状況認識システムに関する検討	第42回日本コミュニケーション障害学会学術講演会, I-3	2016.5
咽喉マイクとピンマイクの同時集音に基づく多人数会話における発話区間推定	第15回情報科学技術フォーラム, 第2分冊, E-020, pp.149-150	2016.9
身体状況認識システムにおける音イベント検出方法に関する検討	日本音響学会2016秋季研究発表会講演論文集, 2-6-2	2016.9
咽喉マイクとピンマイクの同時集音に基づく多人数会話における発話区間推定に関する研究	電子情報通信学会技術研究報告, SP2016-43, Vol. 116, No.279, pp.15-20	2016.10
Voice Activity Detection Using Throat and Lavalier Microphones for Multi-Party Conversations	NCSP2017, 2AM2-3-2, pp.369-372	2017.3
咽喉音を利用した会話・摂食行動の認識	DICOMO2017 シンポジウム, 1D-5, pp.116-123	2017.6