

非侵襲簡易型身体状況認識システムに関する研究*

○安藤純平, 西田昌史, 西村雅史 (静岡大・情報)

1 はじめに

高齢者が健康を維持する上で「食べること」と「話すこと」の重要性が指摘されている。前者は嚥下機能によって支えられており、また、後者は認知の機能と関連が深い。嚥下機能の低下は誤嚥、さらには肺炎等の病気につながり、健康に重大な影響を及ぼすことになる。また、人との会話は認知機能の維持に重要な要素となっている。

我々は簡便なシステムで咽喉音や環境音を収録し、それらを活用することで高齢者の「食べること」「話すこと」に関わる身体状況を把握する方法について検討を行ってきた^[1-3]。今回、日常生活を模した状況下で音データを新たに収録し、嚥下、咀嚼および発話を対象とした識別性能の評価を行った。

2 システム構成

Fig. 1 に示すような集音マイク (Sony ECM-CS3) および咽喉マイク (南豆無線電機 SH12-iK) を首に装着し、小型の MP3 レコーダーで音をステレオ収録する。予備実験として後期高齢者に対して 24 時間装着を依頼し、特に支障がなかったこと、また、目的とする音が問題なく収録されることを確認した。また、若年者を対象とした個別行動の咽喉マイク収録に対し、咀嚼、嚥下、発話などの識別が可能であることを確認している^[1]。今回はより実環境に近いデータを対象とし、Fig. 2 に示すような行動識別システムを構成した。

収録音は咽喉マイクで収録された咽喉音と集音マイクで収録された環境音で構成される。はじめに咽喉音を用いてパワーベースの VAD によって音イベント単位で検出する。その後音イベント単位で咽喉音および環境音から特徴抽出し、各イベントを識別器で嚥下、咀嚼、発話およびその他の雑音の 4 種類に識別する。本来誤嚥と関連が深いとされる喀痰

についても認識対象と設定するべきだが、十分な量の対象データを得ることができなかったため、今回は認識対象から除外した。

特徴量としては MFCC とエネルギーの各イベント内平均値を用いる。また、各イベントの継続長、咽喉音と環境音のパワーおよびその比の各イベント内平均値、環境音の MFCC とエネルギーを特徴量に追加して識別性能の改善を図る。“咽喉音と環境音のパワー比”を特徴量に検討した理由として、嚥下音は集音マイクに収録されず咽喉マイクにのみ収録され、発話は両方で収録されることが挙げられる。なお、予備実験として咽喉音と環境音のパワー比を用いてフレーム単位で嚥下検出実験を実施し、その有効性を確認している。

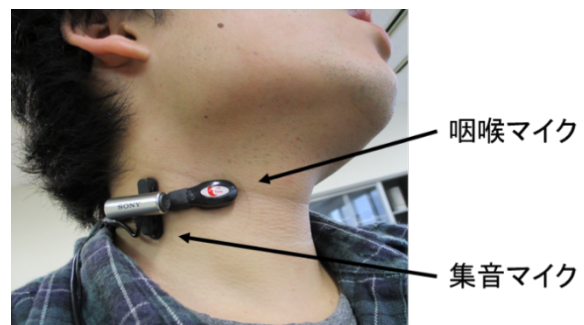


Fig. 1 本研究で用いた咽喉マイクおよび集音マイクとその装着例

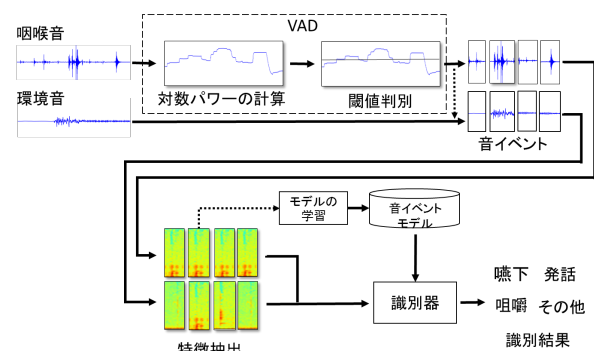


Fig. 2 識別処理の流れ

* A Study on Physical Condition Recognition System Using Non-Invasive Devices, by J. Ando, M. Nishida and M. Nishimura (Shizuoka University).

3 評価実験

3.1 データ収録

日常生活を模した設定で収録した音を用いて評価を行った。被験者は20代の男子学生、計6名で各被験者が10分程度の音収録を行い、合計1時間程度の音データを収録した。被験者には食べ物（ポテトチップス）と飲み物を提供し、これらの摂取をお願いするとともに、静かな部屋で読書もしくはスマートフォンを操作して過ごしてもらう。また、発話を促すため、2分間程度、実験実施者が口頭で質問を行った。収録データに手動で嚙下、咀嚼、発話のラベルを付与し、認識対象ではない咳払いやマイクと服や手との接触音に関してはその他雑音としてラベルを付与した。

3.2 実験方法

識別器には決定木、GMM(混合数 2)、AdaBoost(弱識別器に決定木を利用)、RandomForest および SVM(ガウシアンカーネル)を用い、咽喉音の MFCC とエネルギーの各イベント内平均値のみ特徴量に用いた場合(ベースライン)の識別性能を比較した。また、特徴量として、各イベントの継続長、咽喉音および環境音の対数パワー、さらにその比のイベント内平均値、環境音の MFCC とエネルギーの各イベント内平均値を追加した場合との比較も行った。

評価は被験者ごとに交差検証を実施し、VAD で検出された音イベントに対する識別性能として F 値を算出した。なお、手動で付与したラベルの最長重複区間のラベルを VAD で検出された各イベントに正解ラベルとして付与している。VAD で検出された音イ

ベントの合計は嚙下、咀嚼、発話およびその他雑音それぞれ 187 個、412 個、527 個および 493 個であった。

3.3 実験結果および考察

ベースラインおよび各特徴量を追加した場合の F 値を Table 1 に示す。識別性能が一番高かった SVM の場合、ベースラインと比べて嚙下 0.03、咀嚼 0.32、発話 0.07 の F 値が改善した。とりわけ咀嚼に関して大幅に性能が改善したが、今回の実験では咀嚼音が咽喉マイクだけでなく集音マイクでも収録されたためだと考えられる。

4 おわりに

今回、日常生活を模した状況下で2種類のマイクを用いた収録による識別性能の改善を確認することができた。今後はデータを増やすと共に、より識別に有効な特性を模索し識別性能の向上を目指す。

参考文献

- [1] 西村他, "生体音と環境音の同時収録による高齢者の行動および身体状態認識に関する検討," 音響学会講演論文集, 2-4-9, pp. 1309-1310, 2015.
- [2] 大高他, "咽喉マイクを利用した多人数会話における発話区間推定," 第13回情報学ワークショップ WiNF2015 論文集, pp. 104-106, 2015.
- [3] 小林他, "嚙下音を用いた水分摂取量推定手法の研究," 情報処理学会論文誌, Vol. 57, No. 2, 2016 (掲載予定).

Table 1 分類手法および特徴量別の平均 F 値

分類手法	特徴量	嚙下	咀嚼	発話	その他
DecisionTree	ベースライン(咽喉音 MFCC_E)	0.59	0.43	0.72	0.43
GMM	ベースライン(咽喉音 MFCC_E)	0.67	0.49	0.81	0.51
AdaBoost	ベースライン(咽喉音 MFCC_E)	0.70	0.41	0.78	0.48
RandomForest	ベースライン(咽喉音 MFCC_E)	0.71	0.50	0.81	0.53
SVM	ベースライン(咽喉音 MFCC_E)	0.78	0.45	0.77	0.56
	+継続長	0.78	0.47	0.77	0.56
	+パワー比, 対数パワー	0.79	0.54	0.82	0.58
	+環境音 MFCC_E	0.81	0.75	0.83	0.63
	+パワー比, 対数パワー, 環境音 MFCC_E	0.81	0.77	0.84	0.64