

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 13 日現在

機関番号：37119

研究種目：挑戦的萌芽研究

研究期間：2014～2016

課題番号：26560265

研究課題名（和文）音声波形解析による摂食・嚥下障害発症の予知技術確立に関する研究

研究課題名（英文）Study on foresight technology of eating disorder and the dysphagia onset by the analysis on the wave pattern of speech sound.

研究代表者

甲斐 達男（KAI, Tatsuo）

西南女学院大学・保健福祉学部・教授

研究者番号：60331899

交付決定額（研究期間全体）：（直接経費） 2,900,000円

研究成果の概要（和文）：摂食・嚥下障害を音声解析によって診断し予知するための音声マーカーの探索を行った。音声録音条件、音声解析ソフトの選定、解析パラメーターの選定、解析に影響するさまざまな因子を排除もしくは最小化するための検討、解析のための発語「ワード」の選定を行い、実験条件を設定した。探索は、嚥下障害を瞬時改善する施術法を施す前と後での解析、性別・年齢別解析、各種コーパスを使用した性別・年齢別解析、について実施した。結果として、スペクトログラム濃淡の個人差を標準化するための技術開発、および、複雑な模様として出現するスペクトログラムの部分的減衰を数値として捉えるための技術開発が必須であることが分かった。

研究成果の概要（英文）：We conducted a search for "voice sound markers" to diagnose and predict eating disorder/ dysphagia by speech sound analysis. Establishment of voice recording conditions, selection of voice analysis software, selection of appropriate analysis parameters, examination to eliminate or minimize various factors affecting analysis, selection of a speech "word" for analysis were done to set experiment conditions. The search was carried out on (1) analysis before and after applying the treatment method to instantly improve dysphagia, (2) analysis by sex and age, (3) analysis by sex and age using various corpus. As a result, we found that it is essential to develop technologies to standardize the individual differences of spectrogram density and to develop technologies to capture partial attenuation of spectrograms appearing as complex patterns as numerical values.

研究分野：音声学

キーワード：嚥下障害 診断 予知 音声解析

1. 研究開始当初の背景

(1) 摂食・嚥下障害は、特に高齢者において、誤嚥性肺炎を誘起するので留意しなければならない。また、本障害によって、刻み食やミキサー食などによる栄養補給を余儀なくされ、QOLが著しく低下する。誤嚥性肺炎による死亡については、日本人の死亡原因の第4位が肺炎であり、肺炎のうちの約7割が誤嚥性肺炎である。従って、摂食・嚥下障害は、高齢者の誤嚥性肺炎によって死に直結する可能性が高く、大きな社会問題となっている。われわれは、超高齢化社会を目前にして、そこで生じる諸問題のいくつかを取り上げてこれまで研究対象としてきた。摂食・嚥下障害に関しては、嚥下食の開発と併行して、嚥下障害の早期発見・早期治療のための研究を実施することにした。

(2) 音声を解析することで発症を予測できるようにになれば、痛みを伴うことなく、かつ、検査が容易であるので多くの国民に安価に検診を受けてもらえ、問題が予見できるようになれば、発症の予防措置をとることが可能となる。現在、約80万人が発症しているとされる患者数を激減させるとともに、患者のQOLの向上を図ることができる。

2. 研究の目的

(1) 本研究では、人の音声スペクトログラム(声紋)などの音声パラメータ上に、摂食・嚥下障害の兆候を示す特異的パターン(音声マーカーとよぶ)を見つけ出すことを最大の研究目的とする。その音声マーカーの有無や強度をコンピューター解析することによって、摂食・嚥下障害発症の可能性や発症時期を定量的に診断し、早期発見・早期治療を目指す。

(2) 音声を解析して、摂食・嚥下障害の発症を予知しようという発想は、構音器官(次図の で囲った部分)と摂食・嚥下障害の原因となる「筋肉の弱体化」(図の で囲った部分)が一致する点から着想した。つまり「摂食・嚥下に関わる筋肉の弱体化が起こり始めたら、当然、構音にも何らかの異常が現れることが容易に予想される」ので、それを摂食・嚥下障害の発症を示す音声マーカーとして利用しようとするものである。

3. 研究の方法

(1) 実験条件の設定に関して以下の検討を行った。

音声録音条件の設定

音声解析ソフトの選定

解析パラメータの選定

解析に影響するさまざまな因子の検討

解析のための発語「ワード」の選定

嚥下障害を瞬時改善する施術法の検討

(2) 設定した条件下で次の解析を行った。

施術する前と後における音声解析

性別・年齢別解析

各種コーパスを使用した性別・年齢別解析

4. 研究成果

(1) 音声録音条件の設定と音声解析ソフトの選定

音声解析ソフトは、日本音声学会推薦のAcoustic Core8(以下、Core)とSUGI Speech Analyzer(以下、SUGI)を使用した。録音システムの検討には、次の二つの方法で同一人物による音声データを録り、双方の音声解析ソフトを用いて比較検討を行った。

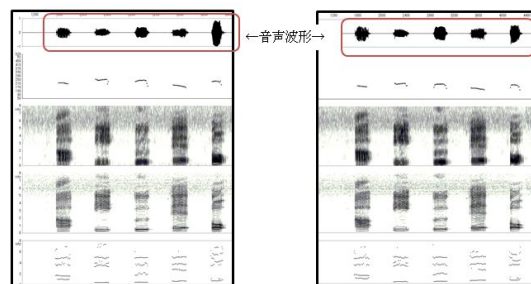
ソコンに外部接続したマイクからインターフェースを介して音声データを直接パソコン内に取り込む。

ICレコーダーによって録音した音声データをパソコン内に取り込む

パソコンのOSはWindows 7 Home Premium、マイクは audio-technica モノラルマイクロホン AT9942、ICレコーダーはSONY ステレオレコーダーICD-SX813を使用した。また(イ)

において、音声データを音声解析ソフトで使用する際、音声データのファイル形式をmp3 から wav へ変換する必要があるため、Sound OrganizerまたはCorel Video Studio ProX4 を用いた。

検討した結果、Coreは解析時間が短く、表示サイズを細かく設定できるが、SUGIは録音時間の長い音声データの解析に優れていることが分かった。本研究の作業効率を考えると、Coreを主たる解析ソフトとして使用し、必要に応じてSUGI、その他の解析ソフトを利用する方針をとることとした。



【左：音声データを直接パソコンに録り込んだ場合、右：ICレコーダーのファイルをパソコンに移した場合】



【左：audio-technica モノラルマイクロホン AT9942、右：SONY ステレオICレコーダーICD-SX813】

より、音声波形において振幅(縦軸)の違いが若干見られた(上図を参照)ものの、他の解析データにはほとんど差異がみられ

ないことから、今後の研究効率を考え、持ち運びが可能である IC レコーダーを使用することとした。さらに、雑音が入らないようにするため、指向性マイクを接続して録音することとした（写真を参照）。

（２）解析パラメーターの選定

Core と SUGI の両ソフトを用いて、次の解析パラメーターの適否を検討した。

音声波形

基本周波数曲線とピッチ曲線

スペクトログラム（広帯域と狭帯域）

フォルマント周波数

実験に用いた音源は次の 8 系統である。

個体差：同日に録音した同年代（20 代）3 人分の音声データを比較

男女差：同日に録音した同年代（20 代）男女の音声データを比較

声の大きさによる差：通常発声とこそこそ声の音声データを比較

口腔による差：通常発声とハミング（口を軽く閉じた声）の音声データを比較

前に発音した音の影響による差：「あ・い・う・え・お」と、発音の順番を入れ替えた「い・え・あ・お・う」の音声データを比較

単語と文章による差：区切って発音した場合と区切らずに発音した場合を比較

アクセントによる差：同じ単語で東京・大阪・鹿児島での方言による音声データ「アメ」を比較

イントネーションによる差：同じ文章で否定・肯定・感嘆の言い方による音声データ「綺麗なない」を比較

尚、ここで用いた音声データは、予備実験のため学術論文や学会などでの公表を目的としたものではなく、倫理審査委員会の審査を受けたものではない。そこで、本報告書の中では、音声波形を示さず結果のみ記載した。

音声波形とピッチ曲線は個体差が大きかった。また、スペクトログラムとフォルマントにおいて、時間（横軸）における若干の固体差がみられるものの周波数（縦軸）は似たような結果が得られた。

音声波形とピッチ曲線は男女で差があった。また、スペクトログラムとフォルマントにおいては、時間（横軸）における若干の違いはみられるものの周波数（縦軸）は似たようなパターンであった。

声の大きさについては、こそこそ声の音声波形が極端に小さく、ピッチ曲線は検出されず、かつ、スペクトログラムの濃度も薄くなった。フォルマントにおいて、SUGI では差がみられたが、Core では差が見られないという結果であった。

口腔の状態による差は、ピッチ曲線ではみられなかった。しかし、音声波形ではハミングの方は小さく、スペクトログラムの濃度も薄くなり、フォルマント周波数でも違いが見られた。

前に発音した音の影響については、音声波形とピッチ曲線に差が出たが、スペクトログラムとフォルマントには差がなかった。したがって、スペクトログラムとフォルマントについては、通常の会話から音を拾って解析できることが推定された。

区切らずに発音するとすべてのパラメーターにおいて異なったパターンが得られたので、区切って発音した音源を解析すべきだと判断した。

「アメ」の発語を、方言で検証した結果、

音声波形とピッチ曲線が 3 つの方言で異なっていた。東京は「ア」が高く、「メ」は低い発音となる。大阪は「ア」が低く、「メ」は高くなるが母音のエから再び低くなる発音である。鹿児島は「ア」が低く、「メ」は高い発音である。

イントネーションによる違いは、「綺麗なない」の発語で検証した結果、音声波形とピッチ曲線で差があった。否定の声は全体的に低く、「な」で高くなり、「い」で低い発音である。肯定の声は全体的に否定より高く、「な」では高くなり、「い」で高くなる発音である。感嘆の声は全体的に最も高く、「な」が高く、「あ」「い」にかけて高くなる発音である。

以上の結果をひとつの表にまとめると次表のように示される。

【発音条件によるパラメーターの違い】

	音声波形	ピッチ曲線	スペクトログラム (声紋)	フォルマント 周波数
個体差	●	●		
男女差	●	●		
声の大きさ	●	●	●	▲
口腔の状態	●	●	●	●
前の発音による影響	●	●		
単語と文章	●	●	●	●
アクセント	●	●		
イントネーション	●	●		

※変化が明らかに見られたものを●で示した。

フォルマントにおける SUGI と Core による差が発生した理由は不明である。声の大小は解析に影響を与える。区切らずに発音をするアクセントやイントネーションによる影響が発現すると考えられた。アクセントやイントネーションによる影響は、音声波形とピッチ曲線で観られた。

以上のことから、音声波形とピッチ曲線は影響を受けやすい因子が多く、スペクトログラムとフォルマントは影響を受けにくいと言える。したがって、本実験での解析パラメーターはスペクトログラムとフォルマント周波数を用いることとした。

(3) 解析に影響する因子の検討

録音を1日のいつすべきか、日間差はないか、喉の調子や、鼻の調子によって差が出ないかについて、次の条件で実験を行った。

録音した時間による差：朝・昼・夜の時間帯に分けて音声データを比較

録音した日による差：3日分の音声データを比較

喉の調子による差：喉の調子が良好な場合と不良の場合の音声データを比較

鼻の調子による差：通常発声と鼻をつまんだ時(鼻声)の音声データを比較

スペクトログラムでは朝、色が薄く、フォルマントでもバラつきがみられるという結果が得られた。つまり、起床して間もなく、発声をあまりしていない朝の録音は不向きであり、録音した時間による影響を受けると考えられる。

録音した日の違いによる影響は、スペクトログラムとフォルマントについては受けにくいと考えられた。

喉の状態は、スペクトログラムでは色の濃淡の違いを発現し、フォルマントでもバラつきが観られた。『日本人の声』(鈴木松美編著)によれば、「喉に炎症を起こした場合などは、声帯や声道の状態も変わり摩擦音を発生することがある」とあることから、喉の調子の違いによって解析に大きな影響が出ると考えられる。

鼻をつまんだ場合の声のスペクトログラムは、濃度がやや薄くなり、フォルマントもバラつきが見られるという結果が得られた。つまり、風邪などによって鼻声しか発音できない場合は音声解析に大きな影響を与えうる。

以上の結果より、録音条件を次のように設定した。

録音時間は昼から夜の間に1回のみ行う。

同一人物による録音は1日のみ行う。

録音は体調が良好な場合に行う。

(4) 解析のための発語「ワード」の選定

無数ともいえるワードの中から、本研究に適した音声解析のための発語(ワード)を決定しなければならない。まず、「有声音」と「無声音」の違いであるが、これは、声帯を使っているか否かであって、声帯は嚥下に直接関係はないため、「有声音」と「無声音」は用いないことにした。次に、「母音」と「子音」についてであるが、母音は、舌の前後の動き・舌の上下の動き・唇の形(円唇か非円唇か)の3つの違いで決まる。一方、子音は、声帯の振動様式・構音点・構音方法の3つの違いで決まる。構音点は、構音器官である舌と軟口蓋の動きが重要であるが、母音、子音ともに舌は発声にも嚥下にも関係している

ため、音声解析ワードの候補とすべきである。構音点は、子音の調整に必要な狭窄や閉鎖がどこに生じるかを示し、「両唇音」「歯音・歯茎音」「歯茎硬口蓋音」「硬口蓋音」「両唇・軟口蓋音」「軟口蓋音」「口蓋垂音」「声紋音」の8種類に分けられる。

『音声・声紋鑑定』(柘植佑好著、立花書房)によれば、「スペクトログラムでの観察で、摩擦音は時間的周期性がなく、周波数成分が調波構造ではないので不規則な強さの濃淡として現れる」「広帯域でも狭帯域でも周期的な縞とならない」とあることから、周期的でなければ観察比較が難しい摩擦音は解析ワードから除外とした。また、摩擦音の構音方法を汲む破擦音も除外とした。半母音は母音を2つ組み合わせており、他の子音に比べ特殊であるため解析ワードから除外とした。八幡総合病院歯科・尾崎由衛先生より助言を頂き、「口腔中を空気が通ることは舌・軟口蓋・喉頭蓋の働きがあり嚥下に関わるため、口腔中を空気が通らず鼻腔だけで音が出る鼻音は嚥下障害の有無にかかわらず変化がないのではないか」という考えから、鼻音をコントロールワードの候補とし、口腔中を空気が通る破裂音と弾き音を嚥下障害改善ワードの候補とすることにした。

解析ワードを最終的に決定するに当たり、「パタカラ体操」と呼ばれる口の代表的な体操を参考にした。「パタカラ体操」とは、「パ」

「タ」「カ」「ラ」という発音によって食べ物を上手にのどの奥まで運ぶ一連の動作を鍛える運動で、嚥下行動にも大きく関わることから介護施設などで普及している体操である。前述した解析用ワードとしての条件に適することから、これら4つを解析用の発語(ワード)の候補とした。これを受け、コントロールワードは、「パ」「タ」「カ」「ラ」の構音点に対応した鼻音と母音を用いることとし、「マ」「ナ」「ガ」をコントロールワードとした。また、「ガ」はそのままでは発音しにくいことから「ッガ」と発音することも可とした。

(5) 嚥下障害を瞬時改善する施術法の検討

同一人物において、通常の音声と、嚙下行動を改善させた状態での音声とを、比較検討することによって、嚙下障害が生じている際の音声パラメーター上の特徴(=音声マーカー)を掴むことを試みることにした。そのための、嚙下行動改善方法として、一般に老人福祉施設や病院で利用されている「アイスマッサージ」と「舌運動」の二つの施術法を用いることとした。

(6) 施術する前と後における音声解析

嚙下障害改善ワード「パ」「タ」「カ」「ラ」の有効性を「マ」「ナ」「ッガ」をコントロールワードとして調べるとともに、上述の施術の有効性の検討を行った。発声は、連続発声「マー・ナー・ッガー・パー・ター・カー・ラー」および、個語の4連発声「マー・マー・マー・マー」(以下、6音も同様)の音声データを比較することとした。20代、50代、70代以降の3世代においてそれぞれ女性3人を被験者とした。録音の順番は次のように行った。施術前の音声の録音、アイスマッサージを施して音声の録音、舌運動を施して音声の録音。被験者は何度か練習し一連の流れを理解した上で実施した。また、解析ワードを書いた紙を用意し、被験者が発話し易いようにした。

尚、ここでも、用いた音声データは、音声波形を示さず結果のみ報告する。

コントロールワード「マ」「ナ」「ッガ」では、連続発声においても、個語の4連発声においても、施術前後による変化は観察されないという結果が全被験者から得られた。一方、嚙下障害改善ワード「パ」「タ」「カ」「ラ」の施術前後による変化は次のような結果が得られた。

ある20代女性では、アイスマッサージ後「パ」の広帯域スペクトログラムが鮮明になった。また、「ラ」では、広帯域・狭帯域スペクトログラムが鮮明になった。また、ある40代女性では、アイスマッサージ後、「カ」の広帯域・狭帯域スペクトログラムが鮮明になった。また、「ラ」と「タ」の広帯域スペクトログラムが鮮明になった。ある70代女性では、アイスマッサージ後、「パ」の広域スペクトログラムが鮮明になった、「タ」と「カ」は、広帯域・狭帯域スペクトログラムが鮮明となった。「ラ」は、狭帯域スペクトログラムが鮮明となった。舌運動後、「パ」は、広帯域スペクトログラムがやや鮮明になり、「カ」は、広帯域・狭帯域スペクトログラムがやや鮮明になった。以上の結果を表にまとめると次表のようになる。

以上のことから、老若間では20代よりも50代・70代の方が変化は大きく、施術前に比べ、アイスマッサージ後・舌運動後共に変

			パ		タ		カ		ラ		
			広	狭	広	狭	広	狭	広	狭	
A	20代	冷香	●							●	●
B	50代	冷香			●		●	●	●		
C	70代	冷香	●		▲		▲	▲	▲		●
			●				▲	▲			

※変化が明らかに見られたものを●、やや見られたものを▲で示す

※「冷」は、アイスマッサージのことを示す

化が見られた。つまり、アイスマッサージと舌運動によって嚙下障害を瞬間的に改善する施術は有効であると言える。また、施術前後の変化をとらえることができた解析パラメーターであるスペクトログラムの有効性も再確認できた。したがって、嚙下障害改善ワード「パ」「タ」「カ」「ラ」は有効であると言える。さらに、嚙下障害改善ワードでのスペクトログラムの有効性が再確認されたことから、施術前後で差が見られなかったコントロールワード「マ」「ナ」「ッガ」も有効であると言える。

連続発声においても、個語の4連発声においても、結果が異なることはなかったため、連続発声のみを行うのがよい。施術前後ではアイスマッサージ後、パターンが鮮明になる場合が多かったため、被験者の負担などを考え、アイスマッサージによる施術のみを行うとよい。高齢者ほど被験者の練習回数が多いと、疲れのためか本番で言いづらくなることがあるため、様子をみながら実施する必要がある。解析に当たり、スペクトログラム(広帯域スペクトログラムと狭帯域スペクトログラムに分かれていない2つが重なった本来の声紋の状態)による比較を取り入れるべきである。

(7) 性別・年齢別解析

20代、50代、70代の3世代において、男女3人ずつの計18名の被験者の音声データを比較解析した。ここでは、SUGIとCore以外にSpectralLayers Proも使用した。

尚、ここでも、用いた音声データは、予備実験のため学術論文や学会などでの公表を目的としたものではなく、倫理審査委員会の審査を受けたものではない。そこで、本報告書の中では、音声波形等は示さず、結果の要点を示すこととした。

SL(Spectra Layers Pro)での解析より、コントロールワード「マ」「ナ」「ッガ」でも、施術前後による変化は観察されないという結果がほとんどの被験者から得られた。一方、嚙下障害改善ワード「パ」「タ」「カ」「ラ」の施術前後による変化も得られた。また、CoreとSUGIでは解析できない8000Hz以上の周波数において施術後鮮明になった部分もあった。また、CoreとSUGIでの解析より、コントロールワード「マ」「ナ」「ッガ」では、

施術後やや鮮明に見える人もいたが、施術前後による変化は観察されないという結果がほとんどの被験者から得られた。

【年齢差・性差・広帯域と狭帯域の割合による変化】

	バ		カ		ク		コ		計
	広帯域	狭帯域	広帯域	狭帯域	広帯域	狭帯域	広帯域	狭帯域	
20代女	1	0	2	0	2	1	2	2	10(42%)
20代男	0	0	1	0	3	0	0	1	5(21%)
50代女	1	0	2	2	1	2	1	1	10(42%)
50代男	2	2	2	2	2	1	2	2	15(63%)
70代女	3	1	2	2	3	3	2	2	18(75%)
70代男	1	0	1	2	2	2	3	2	13(54%)
小計	8(44%)	3(17%)	10(50%)	8(44%)	13(72%)	9(50%)	10(56%)	10(56%)	71(49%)
合計	11(24%)		18(50%)		22(61%)		20(56%)		

上表に示されるように、今回採用した発語(ワード)の解析結果に期待した変化が観られたことから、3世代・男女・複数の被験者から有効性が証明された。さらに、20代と比較して50代・70代の嚥下障害改善ワードの変化数が増加し、男女の変化数の差がほとんどないことから、男女関係のない加齢に伴う変化のデータとして有効と考えられる。『日本人の声』(鈴木松美編著)によれば、「声は25歳を過ぎるころから劣化していく。声帯と、口の構えを作る筋肉、神経伝達能力が劣化するため」とあるが、今回の20代の被験者は全員25歳以下だったため、結果に影響はなかったと考える。今後の実験において、20代の被験者は25歳頃までを目安とする必要がある。また、一個人のスペクトログラムがどのように変化していくのか追跡研究を行うことで、加齢に伴う変化のデータが確実に得られるのではないかと考える。

広帯域スペクトログラムは時間的な変化をとらえることから、広帯域スペクトログラムで変化が見られる=アイスマッサージによる嚥下障害を瞬間的に改善する施術の有効性が高い、と考えられる。

今後の研究において、患者を対象(被験者)とする場合、どの段階で誤嚥を起こしているかを見極めることが重要であろう。嚥下障害はムセ方によって障害されている器官が異なる。口に入れてすぐにムセを起こす場合は口腔(つまり、口腔期)、飲み込む時にムセを起こす場合は咽頭(つまり、咽頭期)、体の状態を倒したことによってムセを起こす場合は食道(つまり、食道期)に問題を抱えている。本実験は口腔・咽頭に関する器官が大きく関わっていることから、口腔期・咽頭期に問題を抱える患者には有効な結果が得られると推測されるが、食道期に問題を抱える患者への結果は不明である。また、顎の角度、身体を倒すなどの姿勢、入れ歯・差し歯などの条件によって口腔内の空気の動きが異なるため、患者にとって避けきれない条件を検討することが必要である。

(8) コーパスを使用した性別・年齢別解析

最後に、加齢に従ってスペクトログラムに有意差が出てくるのかどうかをより多くの音源について確認した。コーパスは、国立情報学研究所の音声資源コンソーシアムが収集した音源データベースの中から、年齢差ごとの解析ができるものとして次の3つのコーパス音源を解析した。鶴岡調査音声データベース91-92(72,600発語の中から解析に適した300発語を解析)、東北大-松下単語音声データベース(3,497発語の中から240発語を解析)、身体情報付き男・女・子どもの母音音声データベース(1,925発語の中から200発語を解析)その結果、音声解析上、克服すべき次の問題点が明らかとなった。

ひとつは、個人によってスペクトログラムの濃淡差が大きく(下図を参照)、これを標準化する技術開発が必要であることであることが分かった。これについては解析ソフト「MATLAB」を用いて濃淡の平均値が最近になって数値化できるようになっているので、標準化技術開発の可能性はある。今後の検討課題である。

もうひとつは、スペクトログラムは大変複雑な模様として出現しているため、大量のデータを解析して一定の法則を見出すには、目視ではなく数値として変化を捉えることが重要となる点である。これについては、画像解析システムである「GTTM分析」が自動で適用できるATTA(Automatic Time-span Tree Analyzer)が最近開発されており、本研究への利用検討が今後の課題である。

5. 主な発表論文等

〔雑誌論文〕(計1件)
準備中

〔学会発表〕(計1件)
準備中

〔その他〕

<http://www3.seinan-jo.ac.jp/university/account/kai/>

6. 研究組織

(1) 研究代表者

甲斐 達男 (KAI, Tatsuo)

西南女学院大学・保健福祉学部栄養学科・教授

研究者番号：60331899

(2) 研究分担者

石本 祐子 (ISHIMOTO, Yuko)

西南女学院大学・保健福祉学部栄養学科・助手

研究者番号：10441867