

認知症患者のQoLを推定するための笑顔と発話の認識

迫 匠一郎[†] 近藤 一晃^{††} 中村 裕一^{††} 松岡 義明^{†††} 大塚 智丈^{††††}

^{††} 京都大学 学術情報メディアセンター 〒 606-8501 京都市左京区 吉田本町

[†] 京都大学大学院 工学研究科 〒 606-8501 京都市左京区 吉田本町

^{†††} 藍野大学 〒 567-0012 大阪府茨木市 東太田

^{††††} 西香川病院 〒 767-0003 香川県三豊市 高瀬町

E-mail: [†]{sako,kondo}@ccm.media.kyoto-u.ac.jp, ^{††}yuichi@media.kyoto-u.ac.jp,

^{††††}{yoshi040905,nishikagawa-ootuka}@yahoo.co.jp

あらまし 本研究では、認知症患者のケアを目的とし、生活の質 (Quality of Life) を表情から推定するシステムの実現を目指している。認知症が高度に進行するまでは、認知症患者の表情表出が健常者と大きく異なることはないため、一般的な表情認識技術の適用が可能であると考えられる。その一方で、一般的なケアの状況では、喜怒哀楽などの表情変化と発話などによる顔の変化が混在するため、単純に表情認識を適用すると誤認識が起こりやすい。この問題に対し、我々は顔部品の位置の時間的変化を用いて、発話と笑顔の識別を試みた。具体的には、顔の各部位の位置変化を時系列信号とし、経験的モード分解を利用して、表情変化によるゆっくりした変化と発話による速い変化を分離する。その際に、顔の上半分 (目の周辺) と下半分 (口周辺) にそれぞれに特徴的な変化が現れることを利用する。本稿では、そのアイデアと簡単な予備実験の結果を紹介する。

キーワード 認知症, QoL 計測, 表情認識, 発話と表情変化の識別, 経験的モード分解

Detection of Smiling and Talking for Estimating QoL of Dementia Patients

Shoichiro SAKO[†], Kazuaki KONDO^{††}, Yuichi NAKAMURA^{††}, Yoshiaki MATSUOKA^{†††}, and Tomotake

OHTSUKA^{††††}

^{††} Academic Center for Computing and Media Studies, Kyoto University Yoshidahonmachi, Sakyo, Kyoto, 606-8501 Japan

[†] Graduate School of Engineering, Kyoto University Yoshidahonmachi, Sakyo, Kyoto, 606-8501 Japan

^{†††} Aino University Higashiohta, Ibaraki, Osaka, 567-0012 Japan

^{††††} Nishikagawa Hospital, Takasecho, Mitoyo, Kagawa, 767-0003 Japan

E-mail: [†]{sako,kondo}@ccm.media.kyoto-u.ac.jp, ^{††}yuichi@media.kyoto-u.ac.jp,

^{††††}{yoshi040905,nishikagawa-ootuka}@yahoo.co.jp

Abstract The purpose of this research is to realize a system that automatically estimates Quality of Life (QOL) of a dementia patient's by monitoring facial expressions of the patient. Because there is no major difference between dementia patients and healthy people in usual facial expressions, we can expect existing facial expression recognition technology can be applied. However, such recognition is often difficult in speaking situations in which facial parts movements occur by talking actions. For this purpose, we explored the classification of position changes, *i.e.*, displacements of facial parts caused by facial expressions or speech. We applied empirical mode decomposition (EMD) to discriminate between slow changes caused by facial expression changes and fast changes caused by speech actions. The upper half of a face and the lower half of face have different characteristics in their changes, and they can be used for discrimination. In this paper, we report the base idea of our method and the results of preliminary experiments.

Key words dementia, QoL measuring, facial expression recognition, discrimination of facial expression and speech, empirical mode decomposition

1. はじめに

高齢化社会が進んだ現在、アルツハイマー病に見られるような認知症の患者数は年々増加し、深刻な社会問題となっている。認知症に対する根治的な治療法は確立されていない現状では、

- 認知症の進行を遅らせる
- 生活の質 (QOL: Quality of Life) を高く維持する

ことが認知症ケアにおいて重要となっている。前者の観点からは、認知症の診断を含め、様々な試みが重点的になされている。その一方で、後者の観点からのケアはまだ遅れているのが現状である。その中で、患者を一人の個人として尊重し、医学的なサポートだけでなく、社会的にも一人の人としてサポートを行う「パーソンセンタードケア (person-centered care)」の考え方が重要視されてきている。このパーソンセンタードケアのためには、患者をよく知ることが必要となる。つまり、それぞれの患者がどのような人であり、何を望み、何ができるか、そのために何を必要としているかを理解することによって初めて、本人の自立的な生活をサポートしたり、さらに、家族やケアセンターでのコミュニティー関係をよりよく維持するサポートをすることが可能となる。

そのための一つの方法として、日常の患者の QOL を評価し、それを治療や介護の指針として用いる動きがある。QOL を指標化することによって、本人の状態や特徴を客観的に理解するとともに、治療や介護の効果を評価したり、好ましくない周辺症状 (BPSD など) が起こる理由を探ることができる。本人による内省や内観が困難な認知症患者の QOL を評価する一つの手法に DCM 法 (Dementia Care Mapping) がある [1]。DCM 法では資格をもった評価者 (マップパー) が患者の様子や周囲との関りや反応を観測し、短い時間 (5 分) ごとに状態の記述とスコア付けをする。これをある程度長い時間 (多くの場合は 6 時間) 連続的に記録することにより、患者本人の状態と性質を知ることが目的とする。しかし、マップパーがつきっきりで QOL を観測することのコストは大きく、観測を頻繁に行うことが難しいという問題がある。

本研究では、上記の問題を部分的に解決するために、自動的に日常の場を観測・記録すること、得られた記録から QOL を評価するための手がかりを得ることを目指している。その方法として、表情認識を用いることとした。現在のセンサー技術やコンピュータビジョンの発展により、そのための基礎的な技術は揃いつつある。その一方で、日常生活下の撮影や表情の認識についてはまだ多くの課題がある。本稿では、このような観点から、自動化のためのシステムをデザインし、その可能性を探った。

2. QOL 計測システム

理想的な自動 QOL 計測システムには以下の要件が考えられる。

- (a) いつどんな場でも患者の状態を記録できることが望ましい。
- (b) 患者の状態を自動的に計測・評価することができる。さらに、計測結果のみを提供するのではなく、医師・介護者・患者の家族に対して現場の理解を深めるための情報を提供できることが望ましい。
- (c) 患者や介護者の普段の行動に影響や支障を与えるもので

あってはいけない。

- (d) 患者や介護者のプライバシーに配慮する必要がある。

まず、(a), (b), (c) より、映像による観測や記録が良い手段となることが想定される。映像による計測は非接触であり、行動を妨げることが少ないことに加え、現場を理解するための良い情報を提供する手段ともなるからである。しかし、上記の項目は各々が難しい問題であり、現在の映像処理技術や環境で各項目を同時に満たすことはできない。そのため、段階的な実現を目指し、以下のように検討を行った。

まず、日常的な認知症ケア現場の環境に適した映像撮影方法や計測システムを設計しなければいけない。その最初の計測対象として、我々は患者が一日の多くの時間を過ごしているとされるケアセンターやリビングルームにあるテーブルを選んだ。(c), (d) の条件より、患者の日々の行動に影響を与えず、また患者が観察されているという不安感を負わせないような撮影方法をデザインする必要がある。また、患者の周辺状況を記録するためには広い視野角の映像が必要になり、これらについては 4 章で述べる。

次に、(b) の計測方法として表情認識を用いることにした。高度に病状が進行していなければ、表情として本人の状態が現れやすいこと、また、それが健常者の表情と大きく異なることによる [3, 4]。図 1 は普段から認知症患者に接している看護師に患者のふるまいを模擬してもらった例であり、介護の熟練者の模擬では表情変化に不自然な違いは見られず、例えば、患者が発話をしながら笑っている表情のように見える。特に、笑顔の頻度と QOL には強い相関があることから [2]、我々はまず笑顔の認識から手がけることにした。この際に重要となる問題として、日常環境では表情変化と発話がよく同時に現れることを考慮する必要がある。この問題については 5 章で述べる。

さらに、(d) に関する別の問題として、患者自身や介護者の顔が撮影・記録されたり他人に見られることに対して抵抗や不安感を与えてしまうことがある。この問題に対しては、CG で作り直した顔に実際に計測した表情を付加するといった表示方法が考えられる。図 2 は QOL 計測結果から、プライバシーに配慮しつつ、介護者や家族に認知症患者の情報を提供するために想定されるユーザーインターフェイスの一例である。これはアイディアの段階であり、本稿ではこれについての詳しい議論は行わない。

3. 関連研究

介護施設における患者の行動の計測や移動の追跡など、認知症患者を観測する手法はこれまでも多数提案されている。中でも、映像で患者を観測し、患者の転倒を検出する研究が多く行われている [5-7]。他にも、高齢者を対象とした行動認識の手法についても多く報告されている [8-10]。このような取り組みは事故を未然に防いだり、ケア環境を改善したり、老化に関連した病気の早期発見を補助することに役立つ。こうした種々の取り組みにも関わらず、QOL の自動計測に関してはまだ方法論がない状況である。まだ QOL として何を計測すべきかという議論が十分に行われていない一方で、介護に従事する人々からは、患者の感情や興



図1 代表的なケア現場の再現

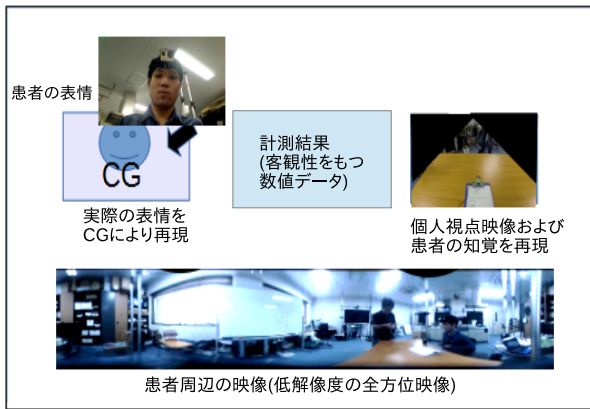


図2 ユーザーインターフェイスの一例

味、注意などを観測したり、それによって QOL の客観的な評価を行うことが期待されている。

また、画像による表情認識の問題に関しては、従来から多くの研究がなされ、良い成果があげられてきた。例えば、顔の向きや表情が変化に対して頑健な顔検出や追跡が提案されている [11, 12]。表情に関しても、顔面筋に基づいて顔の基礎的な動きを効率よく現す Facial Action Coding System (FACS) など [13]、表情認識のために有効な手法が提案されている。また、RGB 等の輝度・色情報とともに深度も測ることのできる RGB-D カメラも安価に市販され、簡単に利用できる顔検出や表情認識のソフトウェアも提供されている。

ただし、上述したように、日常的な環境では、様々な行動に起因する顔の変化が混在することを考慮する必要がある。本研究で想定する介護施設や在宅環境においても、他人と会話をしたり、歩き回ったり、作業など種々の行動が起こる。その中でも、会話することにより、表情変化に起因する顔の変化と発話に起因する顔の変化が混在することが頻繁に起こる。このような問題に関し、発話の際に表情認識精度が下がる結果が報告されており [14]、この問題に対処しなければならない。

まず、発話区間を取り除く手法があるが [15]、これは本研究の目的に合致しない。発話によって変化しやすい顔の下半分を除き、

顔の上半分だけで表情認識を行う手法も提案されている [16]。この手法では発話中の認識精度は改善されるが、発話のない状態では、却って認識精度が悪くなってしまふ。Datcu や Shah らは、発話中には顔上部の顔特徴点を利用し、発話していない場合には顔全体の特徴点を利用する手法を提案している [17, 18]。Wu らは発話の音韻情報を用いて、発話時の顔 (特徴点の位置) から発話していない顔 (特徴点の位置) を推定し、その差を用いて表情認識を行っている。これらの研究のように、音韻情報が利用できれば、発話時の表情認識の精度を上げることが可能となる。しかし、前節の (a)–(d) で述べたように、発話認識は我々が想定している環境では、環境音が大きかったり、マイクを利用することができない場合など、発話の音韻情報が利用できないことを想定する必要がある。そのため、これらの研究とは異なった検討が必要である。

4. 映像記録

表情を自動認識するためには、顔を一定以上の解像度で撮影することが必要となる。また、被撮影者が動くことを考えて、十分な視野も必要となってくる。この2つの観点から、テーブル付近の表情認識に関して検討を行った。患者が食事、作業療法、回想療法などを行っている場合、患者が動く範囲はそれほど広くないが、テーブルの端から端ぐらまでの範囲は動くことが想定される。

一般的な HD 画質 (1920 x 1080 画素) のカメラを利用した場合、図3左図のように、顔領域に 200x200 画素程度の解像度を持たせた状態で、2m 程度の範囲を撮影できる。SD 画質のカメラであれば、幅 60cm 程度の領域の記録が可能となる。このように、顔の向きの変化を考えなければ、一台のカメラで比較的ゆったりとしたテーブルの周辺をカバーできることになる。

次に顔向き方向、およびカメラ配置について考える。顔が斜め方向を向いたり横を向いたりすると、顔検出や表情認識は困難かつ精度が落ちる。つまり図3真ん中の図に示すように、カメラは顔の正面方向から一定の範囲内にある必要がある。

ここで、顔の正面方向に対するカメラ方向の許容される上限を θ_t とし、顔の回転する上限を θ_a とする。この条件を満たすために必要となるカメラの台数は θ_a/θ_t (端数切り上げ) となる。顔位置が移動する場合には、 $\theta_a/2\theta_t + \delta$ 台 (端数切り上げ) 必要である。ただし、小さな移動であれば、 δ は 0、大きな移動であれば 1 以上の値となる。このように、被撮影者の顔の向きを拘束せずに撮影するためには、通常複数のカメラが必要となる。

我々の予備実験において、Constrained Local Models (CLM) を用いて顔追跡や顔特徴点の検出を行った場合、精度良く顔追跡や顔特徴点の検出を行うためには、 θ_t は 50 度程度となった。ここで、被撮影者の顔の回転範囲を 180 度、移動を 1m 以内とすると、上記は $180/(2 \times 50)$ (端数切り上げ) となり、カメラは 2 台必要である。

5. 表情認識を対象とした特徴量

5.1 対象

本研究では、会話や独り言があり得る状況での表情認識に対し

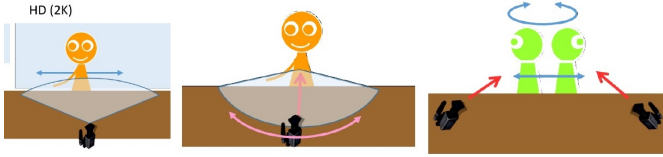


図3 カメラ配置の設計

て以下の条件を想定している．

- 音韻や音声言語の情報が利用できない状況でも利用できること

- 発話と表情変化はそれぞれ独立して生じること

介護施設や在宅の環境を考慮すると、他の人の会話や現場での雑音などの様々な音声情報が混ざり合う状況が想定されるため、正確な患者の音声認識が難しい．また会話の内容から感情を認識することも困難である．人間はそのような状況下であっても、比較的簡単に会話や表情を認識することができる．我々はこのような処理の実現可能性を吟味するために、以下を確認するための予備実験を行った．

- 発話と表情変化が混在しうる状況で、表情変化と発話の有無を認識する

- 発話による顔の変化と表情変化を分離する
- 簡単のため、表情は笑顔に限定した．笑顔は幸福感や充実感との関わりが強いので、QOLを測る良い尺度となるからである．

ここで、発話および笑顔が共にない場合、発話は有るが笑顔でない場合、発話はないが笑顔の場合、発話および笑顔が共にある場合の特徴量を $n_i(t)$, $sp_i(t)$, $sm_i(t)$, $spm_i(t)$ とする．ただし、 i は顔部位の識別子、 t は時刻を表す．特徴量としては、以下では位置の変位量を考える．これまでの先行研究では、実際の特徴量が $sp(t)$ と $sm(t)$ の $lceil$ 線形和 $rfloor$ となることを仮定するものが多いが、線形性を前提とするための根拠は得られていない．

5.2 顔特徴点における実際のパターン

各時刻におけるそれぞれの顔特徴点の変位量は基準点からの距離を基に計算している．基準点の位置を p_0 、発話および笑顔のない場合での基準点から対象とする顔特徴点までの距離を d_i^0 と定義する．また時刻 t における各特徴点の位置を $p_i(t)$ とすると、変位量は

$$d_i(t) = \|p_i(t) - p_0\| - d_i^0 \quad (1)$$

で表される．

実際に用いた特徴点および基準点を図5に示す．顔の下半分の特徴として、口元周辺(左右の口角、上唇の上端、下唇の下端)の4点を選択した．発話と表情変化がない場合のこの4点の重心を仮想的な基準点とし、各時刻の基準点の位置は鼻からの相対位置を用いて推定する．顔の上半分に関しては、左右両側の眉の両端と上端を用いた．基準点は目の中心とする．

発話による顔の動きは発話中の音素、特に母音に大きく依存する．この母音による影響を調査するため、“今日はいいい天気ですね”という文を笑顔あり、笑顔なしの2つの条件で発話したデータを集めた．この文は母音の列にすると“o/o/a/i/i/o/e/i/e/u/e”

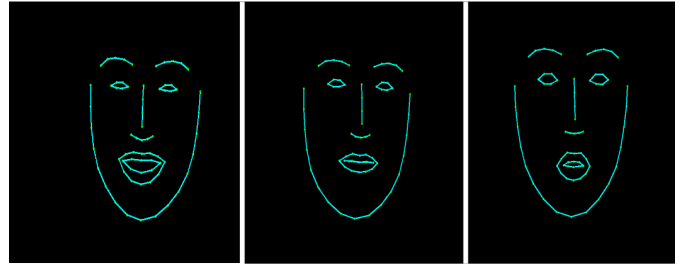
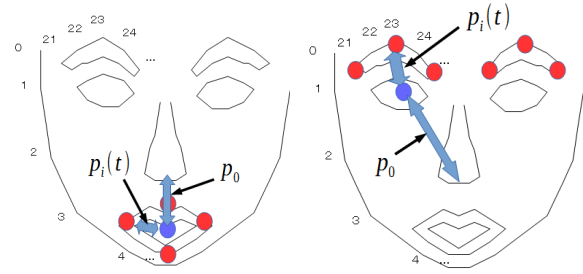


図5 各表情ごとで検出した顔特徴点



(a) 口元周辺の特徴点

(b) 眉部分の特徴点

図6 顔特徴点および変位量の取得

となっており、日本語の全ての母音を含んでいる．また、各母音の発音には 1/5 から 1/2 秒ほど要する (30fps での撮影で 6frame から 15frame) ．

図4の(a)–(e)は上記の口角の典型的な変位量をプロットしたものである．これらのグラフより、笑顔の状態で“a”、“i”、および“e”を発話した場合には、発話のみの場合に比べて変位量が大きくなっていることが分かる．つまり、 $spm_i(t) = k_1 sp_1(t) + k_2 sm_i(t)$ とモデル化すると、 $k_1 > 0$, $0 < k_2 \leq 1$ となる．一方で、“o”または“u”を発話する際には、変位量が減る方向へ変化している．この場合、 $k_1 > 1$, $-1 \leq k_2 < 0$ である．

図4の(f)–(j)は眉頭の典型的な変位量をプロットしたものである．眉に関する変位量は発話の有無に関わらず、表情変化の影響が現れにくい．上記の式を用いると、この場合、 $k_1 \sim 0$, $0 < k_2 < 1$ である．このようなことから、顔の上半分だけで表情を認識することの難しさがわかる．また、 $sm_i(t)$, $spm_i(t)$ の差が小さいため、2つを識別することも難しい．

以上のように、発話と表情変化が混在する場合には、変化をそれぞれの変化の単純な線形和として近似することは難しいことがわかる(注*)

5.3 周波数解析

上記の各特徴点の変位量のパターンに対してフーリエ変換を行った例を示す．図7の(a),(b)はそれぞれ、口角の変位量、眉頭の変位量に対するフーリエ変換結果である．(a),(b)より、発話の有無により周波数特性に違いが出ることがわかる．しかし、口角では、発話のみの時と発話と笑顔が混在している場合を識別する特徴量を設定するのが難しい．眉頭では一見して両者の差が

(注*)ただし、先行研究のように、音声情報から母音が識別できることを仮定すれば、母音ごとに個別の対応が可能である．

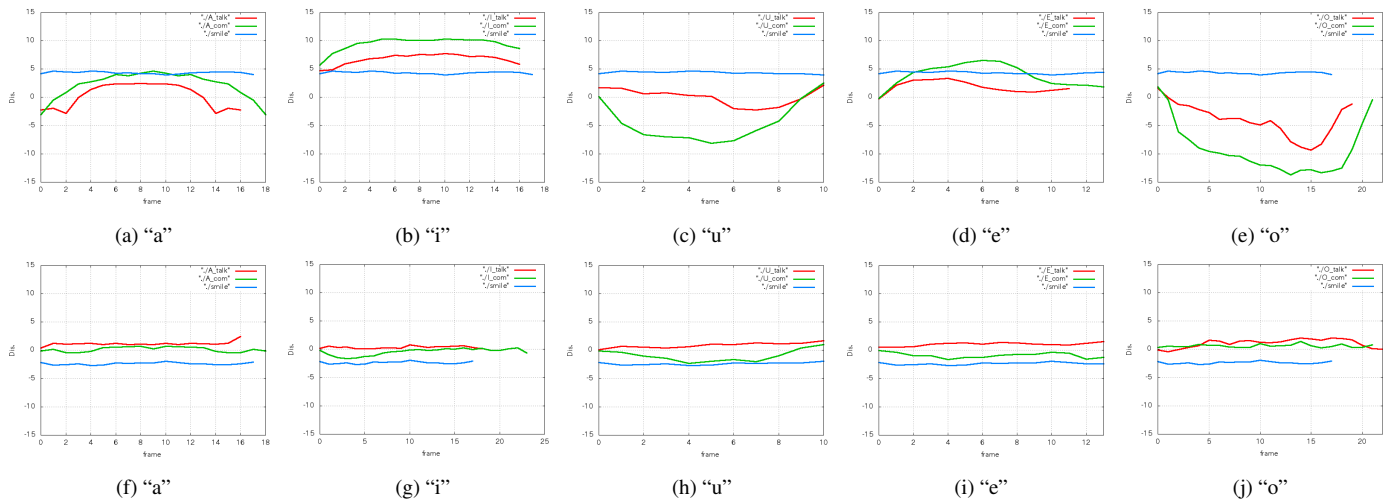


図4 母音発話の際の変位量。上部のグラフは口角の変位量。下部のグラフは眉頭の変位量。
(赤線:発話単体, 青線:笑顔単体, 緑線:混在)

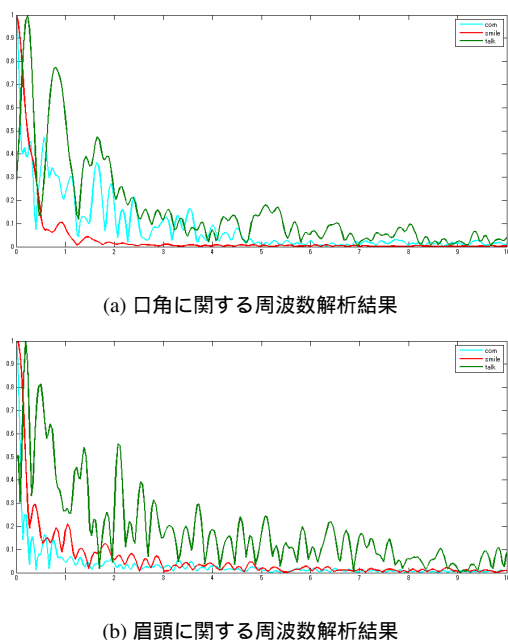


図7 特徴点の変位量に対する周波数解析結果 (赤線:笑顔単体, 緑線:発話単体, 青線:混在)

得られているように見えるが、発話と表情変化が混在する場合に最も低周波成分が多いことなど、説明の困難なものとなっている。これは、図4の下段のグラフからわかるように、実際には小さな差が雑音などに大きく影響された結果であると考えられる。これらの結果は、笑顔による顔面の変化はゆっくりであるため、フーリエ変換のみを用いて表情変化(笑顔)の有無を識別することが困難であることを示唆する。

5.4 経験的モード分解 (EMD)

経験的モード分解 (EMD) は信号を固有モード関数 (IMF) へと分解する手法であり、各 IMF は以下の特徴を持っている。

- 極値の数と零交差点の数は等しい、もしくは多くても差が1である。

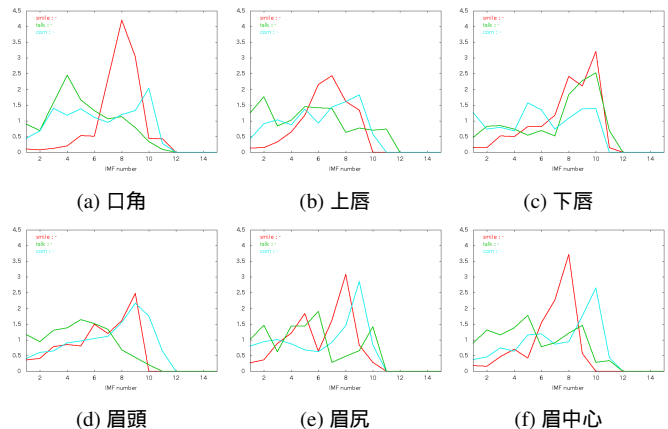
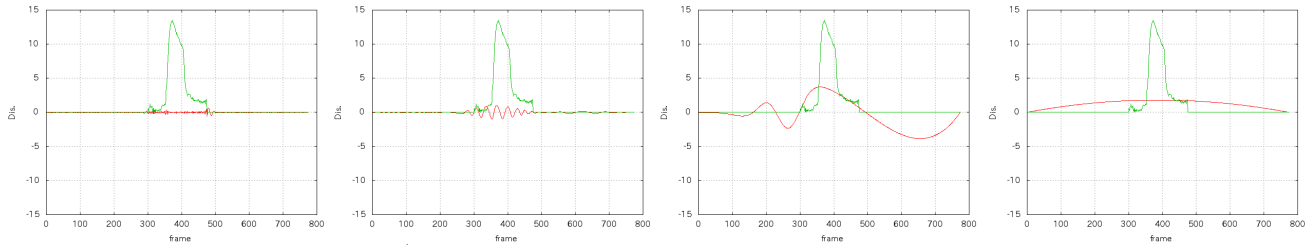


図9 IMFの相対強度。横軸はIMFのオーダー番号。(上グラフ:口元周辺の特徴点, 下グラフ:目元周辺の特徴点, 赤線:笑顔単体, 緑線:発話単体, 青線:混在)

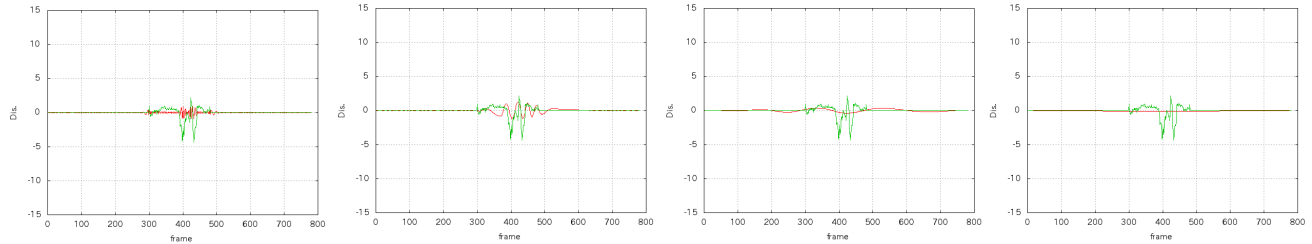
- 極大値と極小値により定義される包括線の平均値は0となる。

各 IMF は原信号と比べ狭帯域信号となっており、また全 IMF と残余信号の合計は原信号となる。小さいオーダーの IMF は通常、原信号での高周波成分を現しており、オーダーが大きくなるほど低周波成分を現す。本研究で EMD を選択した理由は、表情変化に起因する成分と発話に起因する成分の割合が IMF によって大きく異なることを期待したからである。その予想が実際に成立するかどうかを以下で確認する。

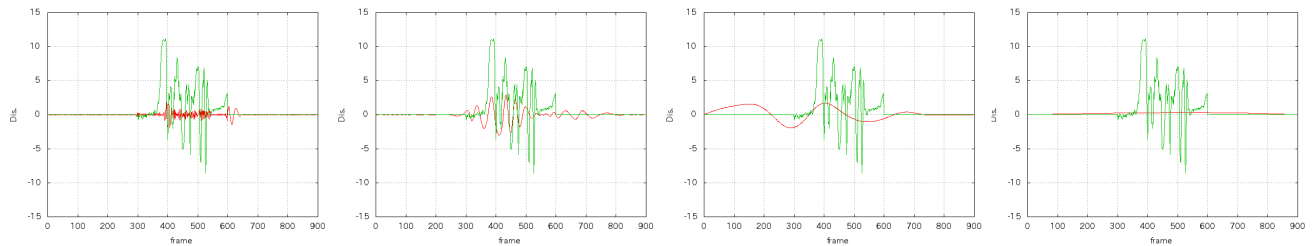
図8は口角の変位量に対して EMD による信号分解を行った結果である。グラフは上から順に $sm_i(t)$, $sp_i(t)$, $spm_i(t)$ に対してそれぞれ EMD を実行した結果を示している。緑線は変位量を示しており、赤線はそのときの IMF のオーダーが 1, 5, 9 および残余信号の結果を示している。グラフから見て取れるように、IMF のオーダーが小さい場合の信号では高周波成分を含んでおり、オーダーが大きくなるにつれて原信号のおおよその形を示す低周波成



(a) 笑顔単体 (IMF number 1, 5, 9, residual)



(b) 発話単体 (IMF number 1, 5, 9, residual)



(c) 混在 (IMF number 1, 5, 9, residual)

図8 口角の変位量に対し EMD により分解した結果 (緑線:原信号, 赤線:EMD により得られた分解信号)

分を含む結果となっている。

各々の結果をこのまま比較することは困難であるため、以下のように各 IMF の相関強度を導出する。まず i 番目の IMF の強度 S_i は積分を用いて

$$S_i = \int IMF_i(t) dt \quad (2)$$

で導出し、相関強度 \bar{S}_i は以下の式で定義する。

$$\bar{S}_i = S_i / \sum_{j=1}^n S_j \quad (3)$$

図9の上側のグラフは、左から順に口角、上唇および下唇に関する各々の相関強度 \bar{S}_i の結果の代表例を示している。まず、 $sp(t)$ では、下唇を除いて、高周波成分において比較的高い数値となっている。また $sm(t)$ では低周波成分において比較的大きな数値となっている。 $spm(t)$ では高周波数帯域と低周波数帯域の両方にピークがある一方、ピーク値は小さいといった特徴を持つ。図9の下側のグラフは左から順に眉頭、眉中心および眉尻に関する各々の相対強度の結果の代表例を示している。眉頭の変位では、 $sp(t)$ で強い高周波成分を持つことが分かり、また $sm(t)$ および $spm(t)$ では強い低周波成分を持つことが分かる。

図8に示すように、IMF が低周波になる(番号が大きくなる)につれて原信号と似た形状をしていることが分かる。つまり、これ

らの IMF は原信号の周波数的大きさおよび時間スケール的な特徴の両方を保持している。そこで、より原信号の特徴を保持している IMF を重視する、つまり重み付けを大きくすることを考える。その重み付けのために、各 IMF および原信号での相互相関を用いる。図10はこの重み付けを行った結果の一例を示している。口周辺(上段)では、 $sm(t)$ において、低周波成分(IMF 番号の大きな部分)に大きなピーク値が見取れる。また $sp(t)$ および $spm(t)$ において大きなピーク値や複数のピークは見られない。眉周辺(下段)では、口周辺と異なるが、それぞれに特徴的な傾向が見られる。

6. 予備実験

EMD を用いた発話および笑顔認識の予備実験を行った。サンプルデータとしてはこれまでの議論と同様、発話のみ、笑顔のみ、混在する場合の3つのデータを集めた。

各顔特徴の変位量を上述した方法で EMD により分解し、サポートベクターマシン (SVM) を適用し、識別の精度を確認した。SVM には、予備テストにより最もデータに対して有効であった線形カーネルを選択した。評価には leave-one-out 法を用いた。

まず、以下のように、(3) 式で定義した各 IMF の相対強度となる特徴ベクトル F を考える。

$$F = (\bar{S}_{1,1}, \bar{S}_{1,2}, \dots, \bar{S}_{i,j}, \dots, \bar{S}_{n,m})$$

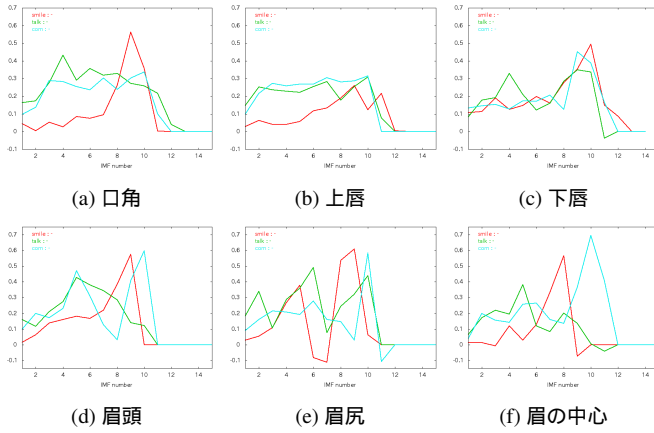


図 10 正規化相互相関で重み付けした IMF の相対強度．横軸は IMF のオーダ番号．(上グラフ:口元周辺の特徴点, 下グラフ:目元周辺の特徴点, 赤線:笑顔単体, 緑線:発話単体, 青線:混在)

表 1 笑顔認識結果 (口元周辺の特徴ベクトル)

入力パターン	識別結果		
	笑顔単体	発話単体	混在
笑顔単体	14 (82%)	0	3
発話単体	0	11 (65%)	6
混在	0	6	11 (65%)

表 2 笑顔認識結果 (目元周辺の特徴ベクトル)

入力パターン	識別結果		
	笑顔単体	発話単体	混在
笑顔単体	11 (65%)	1	5
発話単体	0	10 (59%)	7
混在	2	3	12 (71%)

表 3 笑顔認識結果 (口元 + 目元の特徴ベクトル)

入力パターン	識別結果		
	笑顔単体	発話単体	混在
笑顔単体	15 (89%)	0	2
発話単体	1	11 (65%)	5
混在	2	4	11 (65%)

表 4 笑顔認識結果 (正規化相互相関により重み付けした口元 + 目元の特徴ベクトル)

入力パターン	識別結果		
	笑顔単体	発話単体	混在
笑顔単体	15 (89%)	0	3
発話単体	0	13 (77%)	4
混在	1	4	12 (71%)

ここで $\tilde{s}_{i,j}$ は i 番目の特徴点に対する j 番目の IMF の相対強度である．

さらに, 原信号と各 IMF の正規化相互相関により重み付けした特徴ベクトル F_c を以下のように考える．

$$F_c = (c_{1,1}\tilde{s}_{1,1}, c_{1,2}\tilde{s}_{1,2}, \dots, c_{i,j}\tilde{s}_{i,j}, \dots, c_{n,m}\tilde{s}_{n,m})$$

ここで $c_{i,j}$ は i 番目の特徴点に対する j 番目の IMF と原信号との正規化相互相関の値である．

これら F, F_c を用いて識別実験を行った結果を以下に示す．表 1-3 は F を用いて場合の結果であり, 口元周辺の特徴点のみ, 眉周辺の特徴点のみ, 口元および眉両方の特徴点を用いた各々の場合での識別結果を混同行列を用いて示した結果である．表中の数字は各行列のラベルに属すると識別された実際のサンプル数を示している．識別結果より, 口元周辺の特徴点は発話の影響を受けて精度が低下していることが分かる．それと比較して, 顔の上半分の特徴点は顔の下半分と比べ, 表情認識に対する効果的な情報を含んでおらず, 発話がない状態においても精度が低い結果となった．また顔の上半分および下半分の両方の特徴点を用いることで, 認識精度は向上した．表 4 は F_c を用いた際の結果を示している．表 3 に比べて識別精度が向上する結果となった．

これらの結果より, 各顔部位の変位量が EMD によって分解される特徴量が笑顔, 発話, その混在の識別のために良い特徴となっていることが示唆された．さらに, IMF と原信号との正規化相互相関により重み付けを行うことで, 識別の精度を向上させることも期待できる．

しかし, 本研究は少数のデータに対する予備実験を行った段階であり, 複数人に対する一般的な環境での種々の調査が必要である．また, 怒りや悲しみ, 驚きといった他の感情も患者の状態推定のために重要であり, 今後の課題として, それらの表情の認識についても調査する必要がある．

7. ま と め

本稿では, パーソンセンタードケアの支援を目的とし, 認知症患者の QOL を映像による計測する取り組みについて紹介した．まず, 患者の観察および QOL 計測のために必要な項目について述べ, 普段の患者の環境に合った映像記録の撮影方法や表情認識技術について検討した．次に, 患者の顔を記録し, 笑顔検出を行う上で, 発話による顔面の動きが表情の認識を難しくするため, 発話と笑顔の分離認識について検討を行った．その結果, 顔特徴点の変位量を EMD により分解した信号が笑顔および発話の識別にとって有効な手段であることが示唆された．これらの実験はまだサンプル数が少なく, 予備実験に過ぎない．今後, 提案手法の有効性を確かめるために, 更なる実験が必要である．さらに, 笑顔以外の表情の認識を検討することも今後の重要な課題である．

文 献

- [1] Brooker, D: "Dementia care mapping: a review of the research literature.", The Gerontologist Vol.45, suppl 1, pp.11-18, 2005
- [2] 花沢明俊ら, "認知症老人の笑顔画像計測による情動機能回復評価", 一般社団法人電子情報通信学会信学技報, pp39-42,(2012)
- [3] Lawton, M., Van Haitsma, K., Klapper, J.: "Observed affect in nursing home residents with Alzheimer's disease", The Journals of Gerontology Series B: Psychological Sciences and Social Sciences Vol.51.1, pp.3-14, 1996
- [4] Jansson, L., Norberg, A., Sandman, P.: "Interpreting Facial Expressions in Patients in the Terminal Stage of the Alzheimer Disease", Omega (Westport), vol.26 no.4 pp.309- 324, 1993
- [5] Nait, C., McKenna, S.: "Activity summarisation and fall detection in a supportive home environment", Int. Conf. on Pattern Recognition (ICPR), 2004.
- [6] Foroughi, H., Aski, B., Pourreza, H.: "Intelligent video surveillance

- for monitoring fall detection of older people in home environments,” Computer and Information Technology, 2008. ICCIT 2008. 11th International Conference on. IEEE, 2008
- [7] Sixsmith, A., Johnson, N.: ”Smart sensor to detect the falls of the elderly”, IEEE Pervasive Computing, vol.3, no.2, pp.42-47, 2004.
- [8] Chen, D., Bharucha, A., Wactlar, H.: ”Intelligent video monitoring to improve safety of older persons”, Proc IEEE Eng Med Biol Soc. 2007.
- [9] Crispim-Junior, D., Bremond, F., Joumier, V.: ”A Multi-Sensor Approach for Activity Recognition in Older Patients”, The Second International Conference on Ambient Computing, Applications, Services and Technologies - AMBIENT 2012.
- [10] Lapointe, J., Bouchard, B., Bouchard, J., Potvin, A., Bouzouan, A.: ”Smart Homes for People with Alzheimer’s Disease : Adapting Prompting Strategies to the Patient’s Cognitive Profile”, 5th ACM International Conference on Pervasive Technologies Related to Assistive Environments (PETRA’12), 2012
- [11] Tirkaz, C., Albayrak, S.: ”Face recognition using active appearance model.”, IEEE 17th Signal Processing and Communications Applications Conference, 2009.
- [12] Cristinacce, D., Cootes, T.: ”Feature Detection and Tracking with Constrained Local Models.”, BMVC, Vol.1. No.2. 2006.
- [13] Ekman, P., Friesen, W.: ”Facial Action Coding System: A Technique for the Measurement of Facial Movement”, Consulting Psychologist Press, 1978
- [14] Lin, J., Wu, C., Wei, W.: ”Error weighted semi-coupled hidden Markov model for audio-visual emotion recognition,” IEEE Trans. Multimedia, Vol.14, No.1, pp.142-156, 2012
- [15] Kotsia, I., Buciu, I., Pitas, I.: ”An analysis of facial expression recognition under partial facial image occlusion,” J. Image Vision Comput., Vol.26, No.7, pp. 1052-1067, 2008
- [16] Zor, C., Windeatt, T.: ”Upper Facial Action Unit Recognition”, Advances in Biometrics, Springer, pp.239-248, 2009
- [17] Datcu, D., Rothkrantz, L.: ”Semantic audio-visual data fusion for automatic emotion recognition,” Euromedia, 2008.
- [18] Shar, M. et. al: ”Action Unit Models of Facial Expression of Emotion in the Presence of Speech”, Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII), pp.49-54, 2013
- [19] Wu, C., Wei, W., Lin, J., Lee, W.: ”Speaking Effect Removal on Emotion Recognition From Facial Expressions Based on Eigenface Conversion”, IEEE Trans. Multimedia, Vol.15, No.8, pp.1732-1744, 2013