

SVMによる笑顔度推定技術を用いた音楽療法効果の評価

嶋田 敬士^{1,2,a)} 山田 亨^{3,4,5,b)} 高崎 友香⁴ 野口 祥宏¹ 山崎 郁子^{4,5,6} 福井 和広²

受付日 2013年9月16日, 採録日 2014年9月12日

概要: 病院でのリハビリテーションの一環として行われる取り組みの1つに、音楽療法士が音楽の力を意図的・計画的に利用して、患者の心身障害の回復や機能の改善に役立てる音楽療法がある。従来その効果は、病院や音楽療法士が各々独自に設けた評価基準と介入内容の記録などを通じて質的・量的に評価する方法が試みられてきたが、患者の症状、回復状況や個性が様々なことなどから、客観的で統一的な評価方法を確立することは非常に困難であった。そこで我々は、患者の心身賦活にともなって広く一般的に見られる表情である笑顔に着目し、評価記録用に撮影された音楽療法時の映像データのみから患者の笑顔度を定量化した。さらに、あらかじめ構造化されていた療法内容に着目し、介入の質が異なる場面での笑顔度の違い、それらの経過回数にともなう変化を統計的に検定した。その結果、療法内容と経過にともない統計的に有意な患者の表情変化を確認するとともに、その変化が従来行われてきた主観評価結果とも高い相関を示していることを確認した。

キーワード: サポートベクタマシン, 表情, 笑顔, リハビリテーション, 音楽療法, 多重比較検定

Evaluation of Music Therapy Effect by Use of Smile Intensity Estimation by SVM

KEIJI SHIMADA^{1,2,a)} TORU YAMADA^{3,4,5,b)} YUKA TAKASAKI⁴
YOSHIHIRO NOGUCHI¹ IKUKO YAMAZAKI^{4,5,6} KAZUHIRO FUKUI²

Received: September 16, 2013, Accepted: September 12, 2014

Abstract: Music therapy is one of the rehabilitation program and that is approach to improve physical and mental handicaps of clients using music by therapists. Conventionally, the music therapist evaluated the effect of music therapy with the original policy and intervention details, but it was very difficult to establish the common and objective evaluation policy, because of the variety of the client's symptom, differences of each hospital policy, therapist experiments and so on. Therefore we focused on the smile especially among several facial expressions, which was observed commonly with mind and body activation of clients and quantified the smile intensity from only image sequences for music therapy records without contact and constrain to the clients. Furthermore, we tested statistically the difference of the smile intensity, which depends on the interventions by therapists and progresses of the therapy. As a result, we confirmed statistically significant facial expression differences of client with the therapeutic regimen and progress and there was high correlation between that and subjective evaluation results.

Keywords: support vector machine, facial expression, smile, rehabilitation, music therapy, multi comparison

¹ 旭化成株式会社融合ソリューション研究所
Synergistic Solutions Initiative, Asahi Kasei Corporation,
Atsugi, Kanagawa 243-0021, Japan

² 筑波大学大学院システム情報工学研究科
Graduate School of Systems and Information Engineering,
University of Tsukuba, Tsukuba, Ibaraki 305-8577, Japan

³ 独立行政法人産業技術総合研究所ヒューマンライフテクノロジー
研究部門
Human Technology Research Institute, National Institute
of Advanced Industrial Science and Technology, Tsukuba,
Ibaraki 305-8568, Japan

⁴ 茨城県立医療大学付属病院

Ibaraki Prefectural University of Health and Sciences Hos-
pital, Inashiki, Ibaraki 300-0331, Japan

⁵ 茨城県立医療大学
Ibaraki Prefectural University of Health and Sciences,
Inashiki, Ibaraki 300-0331, Japan

⁶ 国際医療福祉大学大学院医療福祉学研究科
International University of Health and Welfare Graduate
School, Otawara, Tochigi 324-8501, Japan

a) shimada.kb@om.asahi-kasei.co.jp

b) toru.yamada@aist.go.jp

1. はじめに

わが国では、2011年10月1日現在、総人口における65歳以上の割合が23.3%にも達し、今や国民の約4人に1人が高齢者となり、さらに75歳以上人口が全体の11.5%を占めるなど、本格的な超高齢化社会を迎えつつある[1]。そして、この少子高齢化傾向は今後も継続し、2060年には2.5人に1人が65歳以上、4人に1人が75歳以上となり、先進諸国の中でも特に高齢化が急速に進んでいくことが予測されている。そのため、介護・医療・福祉の充実とそれに付随する社会保障費の削減という相反する課題をともに解決することが強く求められている。特に近年は、このような長寿高齢化の社会構造の変化にともなって、病院を訪れる患者の疾病内容や治療方法も変化、多様化してきており、生活習慣病のケア、脳卒中後や要介護者の自立を促すためのリハビリテーションなどの重要性が年々増してきている。このような背景から、近年は絶対的な患者数の増加が一因となり、リハビリテーションが必要な患者に対して適切な方法と期間を持って施術が行えないなどの問題も始めている。この問題を解決するためには、社会的な体制、施設や人員を備えていくことも必要ではあるが、それと同時に医療施設などにおいて患者の疾患から完治、ポストケアに至るまでの治療サイクルをいかにロスを少なく最適化していくかが重要である。そしてそのためには、個々の患者の回復状態を適切に評価・把握しつつ、必要なリハビリテーションを計画的に実施するため、リハビリテーションの効果を客観的・定量的に評価することが臨床現場での重要な課題の1つとなってきている。

そのような病院でのリハビリテーションの一環として行われる取り組みの1つに音楽療法がある。音楽療法とは、「音楽のもつ生理的、心理的、社会的働きを用いて、心身の障害の回復、機能の維持改善、生活の質の向上、行動の変容などに向けて、音楽を意図的、計画的に使用すること」と定義されている[2]。本療法においては従来、患者に対する療法の効果を、病院や音楽療法士が各々独自に設けた評価基準と介入内容の記録などを通じて質的・量的に評価することが試みられてきた。しかし、患者の症状や回復状況が様々であり、また評価者としての音楽療法士の判断基準も経験などによって左右される部分も少なからず存在することから、客観的で統一的な評価方法を確立することは非常に困難であった。そこで我々は、人間の心身賦活にともなって広く一般的に見られる表情変化である笑顔に着目した。

そもそも我々の社会生活において、視覚から得られる情報の果たす役割は非常に大きく、特に人と人とのコミュニケーションにおいては、言語に加え、相手の仕草や顔の表情が視覚的に確認できることで、我々はよりスムーズに、より深く相手の意図を理解することができる。人間の表情の解析については、心理学[3]をはじめとした様々な学術

分野からアプローチがなされているが、工学の分野[4]においても、表情や感情を顔画像、声、生体信号などを用いていかに機械により解析・推定・理解するかという試みが多くの研究者によってなされてきた。しかし、顔画像のみから表情を認識することは、データ収集の難しさ、表情判定の難しさ、収集データと実環境との乖離、表情と感情とのずれなどの問題から、いまだ実環境下で十分な性能を発揮する認識手法が確立されたとはいいがたく、今もなお活発に研究がなされている。また、そのような人間の表情を治療に役立てようとする試みとしては、たとえば様々な表情の合成顔画像から脳障害の程度を評価しようとする試み[5]や、人の情動変化に起因する表情の変化を依存症の診断に役立てようという試み[6]などがある。

そのような人間の表情の中でも特に笑顔は、我々の日常生活においてコミュニケーションや社会活動を円滑に進めるうえで重要な役割を果たす表情の1つであり、また音楽療法中の患者の症状の改善過程にともなっても頻繁に観察されるようになるポジティブな表情変化であることが経験的に分かっている。しかし、このように主観的には笑顔の発生頻度や度合いの変化が音楽療法の効果と何らかの関係がありそうなが分かってはいても、従来はそれを定量化して客観的に評価する術がなかった。そこで我々は、映像データのみから非接触・非拘束に人物の顔を自動的に検出し、その顔画像のみから笑顔度を推定し定量化する技術を応用し、音楽療法セッション記録映像から患者の笑顔度を定量的に測定し、療法経過や介入内容による変化を多重比較することによる音楽療法効果の客観的な評価方法について提案し、その有効性を評価した。

本稿の構成であるが、2章ではリハビリテーションにおける音楽療法の位置付けと表情との関連性について述べ、3章では映像中から笑顔度を推定する手法について説明し、その性能評価結果について4章で述べる。そして、5章で実際の臨床現場で撮影された映像から笑顔度を推定した結果と統計的検定による客観評価結果について報告し、6章でまとめを述べ、最後に7章で今後の展望について言及する。

2. 音楽療法と表情の関連性

2.1 リハビリテーションにおける音楽療法

リハビリテーションとは、疾病や外傷によって心身の機能が障害され、日常生活や社会参加に支障を来している個人が、複数の医療専門職の支援と連携の下で自身の問題を解決していくプロセスととらえることができる。そこで個人が直面する問題はじつに様々であり、また各専門職が支援する分野や用いる手法も異なる。たとえば、理学療法士や作業療法士は歩行や上肢動作の機能回復訓練を支援する。しかし、リハビリテーションの対象のうち脳機能障害患者などは、しばしば覚醒水準や発動性の低下や抑うつを呈し、情動反応も低調である。このため内発的な回復意欲

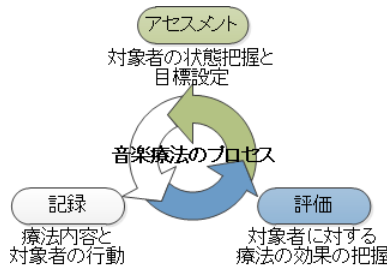


図 1 音楽療法のプロセス

Fig. 1 Process of the music therapy.

に欠け、リハビリテーションの効果が上がりにくい場合がある。こうした症例に対して患者の覚醒水準や発動性の向上を期待して音楽療法が行われている。一般的な音楽療法は図 1 に示すようなプロセスサイクルに従って計画的に実行され [7]、我々は、病院内でこのようなプロセスに従い小集団の音楽療法を行い、患者の情動反応や発動性の向上を促して機能回復訓練へとつなげる取り組みを十数年にわたって行ってきた。

2.2 療法効果の評価

このような取り組みを効果的に進めるためには、覚醒水準、発動性、情動反応のような精神機能面での対象者の変化を異なる複数の医療専門職の間で的確に共有することが重要である。そのため、現状ではあらかじめ用意した主観評価表などを用いた点数化を行っているが、症例によっては評価項目の妥当性や、評価者の熟練度による点数の信頼性などの問題がある。また、作業療法士らが音楽療法の現場に立ち会って担当患者の発動性の改善を直接確認する例も、取り組みの効果が認識されるにつれて増えつつあるが、このような立会いをいつも行うことは難しい。ゆえに、こうした取り組みの効果を広く一般の医療関係者に伝達するためには、異なる症例でも共通に適用可能な客観的評価手法を通じて、その介入効果を検証することが必要である。

2.3 表情の活用

精神機能を客観的に評価する手法は今なおほとんど提案されていない一方で、喜び、悲しみ、怒り、驚き、嫌悪などの情動と強く結び付いて表出される表情は人種や文化を超えて広く共通していることが知られている [3]。そこで我々は、覚醒水準や発動性の向上にともなって患者に現れる情動反応を表情を通じて客観的に評価できれば、こうした精神機能を評価する有力な手段になると考えた。

音楽療法では、同じ音に対して状況によって対象者が異なる情動反応を示すこともしばしばあり、療法士はその反応につねに注意深く向き合いつつ、望ましい行動を最大限に引き出すために介入の質と量を調整しなければならない。表情の中でも笑顔は快感情の表出である点で他の表情とは明確に異なり、この点が介入において非常に大きな手

がかりを与えている。実際、快感情すなわち笑顔を誘起するように介入を調整することによって、対象者に深刻な悪影響を及ぼす事態を回避できるばかりか、望ましい行動を引き出せる見込みは高まる。そのため、ほとんどの音楽療法は不快感情ではなく快感情を誘起するように、またそれが対象者の達成すべき目標行動の獲得に合致するように設計される [7]。また、怒りや悲しみなどの不快感情の低減を評価尺度に用いる場合に比べると研究デザインが単純ですむため、評価結果の解釈に曖昧さが少ないという利点もある。こうした背景から我々は、音楽療法効果の評価尺度として笑顔度が様々な現場で広く利用できると考えた。

3. SVM による笑顔度推定手法

本研究では、対象者に先入観や警戒感を極力いだかせないため、特別な器具や装置の装着を必要とせず非侵襲・非接触・非拘束に笑顔度を推定する方法として、ビデオカメラなどで撮影された映像データのみから対象者の笑顔度を推定することを試みた。一般的に、笑顔を含む表情を映像情報のみから認識する手法は 2 つに大別される。1 つはアピアランス・ベースの手法 [8], [9], [10] で、もう 1 つは特徴点ベース（またはモデル・ベース）の手法 [11] である。アピアランス・ベースの手法は、特徴点ベースの手法に比べ、顔を検出した後に目、口や鼻といった顔器官を検出する必要がないため処理が単純で高速であるが、検出された顔の位置ずれ、顔向きの変化や個人ごとの顔器官配置の違いの影響を受けやすい。一方、顔検出後に顔器官を検出する特徴点ベースの手法は、処理は複雑になるが特徴点周辺の詳細な解析、顔の検出位置ずれや顔向きの補正が可能のため、一般的にはアピアランス・ベースの手法に比べて表情の検出性能が頑健であるが、性能が顔器官の検出性能・精度に依存するなどの問題点もある。本研究では、画像認識技術が広く一般的に利用されるためには、プライバシー保護の観点から映像データをストレージなどに保存することなくリアルタイムに処理することが重要な課題の 1 つと考え、高速かつロバストな笑顔度の推定を目指し、カラー情報を使わずまた顔器官などの特徴点を抽出する必要がないアピアランス・ベースの特徴抽出法をベースに多少の顔の検出位置ずれや顔向き変化にも対応可能な特徴抽出法について検討した。図 2 に本提案手法の全体処理フローを示す。まずはじめに 256 階調グレースケールの入力画像から顔を検出し、検出された顔領域から顔画像を切り出し、スケールと照明の影響を低減させるための正規化を施した後に、アピアランス特徴量を抽出し、笑顔の検出を行いその出力から笑顔度の推定を行う。

3.1 顔の検出

映像から笑顔度を推定する第 1 ステップとして、入力された映像から対象者の顔を検出する必要がある。本研究

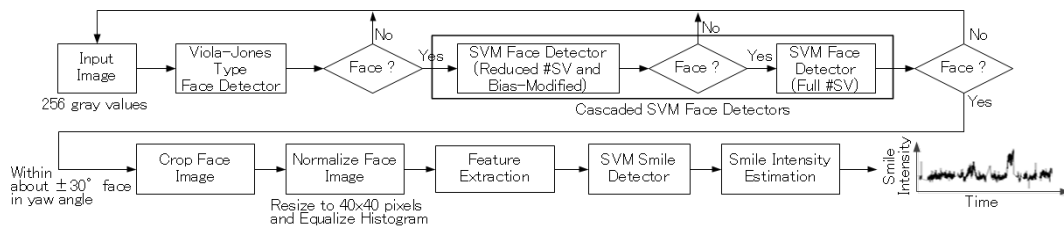


図 2 提案手法の処理フロー

Fig. 2 Process flow of our proposed system.

では、図 2 のフロー上段に示すように、顔検出器として Viola-Jones 型の顔検出器 [12] とカスケード構造化された 2 つの Support Vector Machine (以下, SVM) [13], [14] による顔検出器を組み合わせた、高速かつロバストな顔検出器 [15] を用いた。Viola-Jones 型の顔検出器はカスケードの段数を適当に設定することで、誤検出をある程度許容しつつも高速な顔検出が行える。一方, SVM は識別性能が高いことで知られるが、顔検出のような非線形問題へ対応するためにカーネルトリックを用いた場合、探索数、サポートベクタ数、次元数に従って計算コストが膨大となる。そこで、まずはじめに入力画像全体から Viola-Jones 型の顔検出器により高速に顔候補を絞り込み、検出されたたかだか数個の顔候補に対してのみ SVM 顔検出器により顔検出を実施する。さらに SVM 顔検出器を、Reduced Set Method 法 [16] を用いて SV 数を削減した少数のサポートベクタからなる副顔検出器と全サポートベクタを用いた主顔検出器の 2 段カスケード構造とすることで、SVM 顔検出器の性能を維持しつつさらなる計算速度の向上を図った。また本研究では、学習用画像として水平方向へ ± 30 度程度以内の顔画像を用いることで、真正面以外の多少の顔向き変化に対しても顔検出が可能のように顔検出器を学習させた。そして、この顔検出器により検出された顔の位置・サイズをもとに、入力画像から顔領域のみを切り出し、サイズを 40×40 pixel にスケールし、照明環境の影響を抑えるためにヒストグラムの平滑化を施した後に、後述する笑顔推定のための特徴抽出処理を施す。

3.2 笑顔度の推定

顔検出器によって検出された顔画像に対し、笑顔度を推定するための特徴量を抽出する。本研究では、目、鼻や口などの顔器官を抽出する必要がないアピアランス・ベースの特徴を笑顔推定のための特徴量として抽出し、その特徴量を入力として SVM による笑顔検出器を構築する。そして、その SVM 笑顔検出器からの出力値をシグモイド関数を用いて事後確率に変換することで、笑顔度を算出する。

一般的に顔検出器により検出された顔は、目や口などの顔器官の位置がつねに一致しているという保証はない。そのため、顔のアピアランス情報のみから精度良く笑顔を検出するためには、顔の検出位置ずれなどに対してロバスト

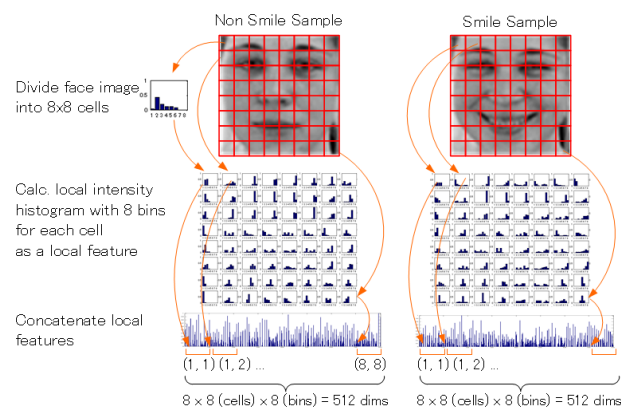


図 3 LIH による特徴抽出例

Fig. 3 Example of feature extraction by LIH (8×8 cells, 8 bins).

な特徴抽出法が必要である。そこで本研究では、顔画像をある一定の大きさの格子セルに分割し、各格子セル内で局所的な特徴量を算出した後に、それら局所特徴量を結合して笑顔特徴量とした。本研究では局所的な特徴抽出法として、局所輝度ヒストグラム (LIH) [17] と Center-Symmetric Local Binary Patterns (CS-LBP) [18] を検討した。以下、各々の算出方法について説明する。

3.2.1 LIH 特徴量

LIH は、局所領域内の輝度ヒストグラムを算出し、それらを連結した特徴量である。その算出ステップを以下に示す。

- (1) 顔画像をある一定の大きさの格子セル ($M \times N$) に分割する。
- (2) 各格子セル領域内で L 階調輝度ヒストグラムを求める。
- (3) 各格子セルの輝度ヒストグラムの総和が 1 になるように正規化を施す。
- (4) 格子セルごとに求められた正規化された局所輝度ヒストグラムを連結し、 $M \times N \times L$ 次元の特徴量を得る。

図 3 に格子セル数を 8×8 、ヒストグラムビン数を 8 としたときの特徴抽出例を示す。

3.2.2 CS-LBP 特徴量

CS-LBP は単純な計算方法ながら、輝度変化に強いテクスチャ特徴を Local Binary Patterns (LBP) [19] よりもコンパクトに表現できることで知られており、近年物体検出などでよく使われる特徴抽出法の 1 つである。その計算方

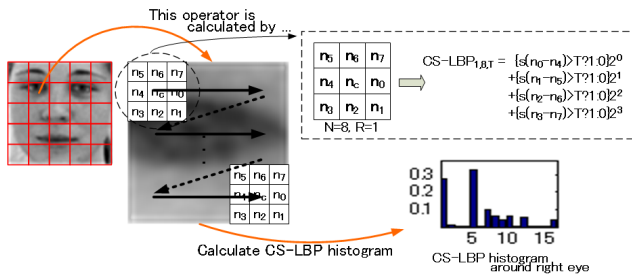


図 4 CS-LBP の計算例

Fig. 4 Example of CS-LBP histogram calculation.

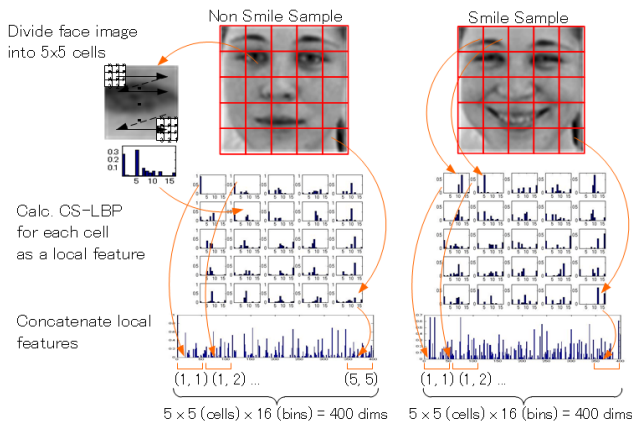


図 5 CS-LBP による特徴抽出例

Fig. 5 Example of feature extraction process by CS-LBP (5 x 5 cells).

法は、注目画素に対して、半径を R 、近傍数を N 、 i 番目の近傍画素の値を n_i 、エンコード閾値を T とすると、式 (1) により求められる。

$$CS-LBP_{R,N,T}(x, y) = \sum_{i=0}^{(N/2)-1} s(n_i - n_{i+(N/2)})2^i$$

$$\text{where } s(x) = \begin{cases} 0 & (x < T) \\ 1 & (\text{otherwise}) \end{cases} \quad (1)$$

ここで本研究では、近傍数 $N = 8$ 、半径 $R = 1$ としたため、注目画素を n_c としたとき、ある注目領域内での CS-LBP は図 4 のように計算できる。

このようにして計算される CS-LBP を用いて、顔画像から笑顔検出のための特徴量を抽出するステップを以下に示す。

- (1) 顔画像をある一定の大きさの格子セル ($M \times N$) に分割する。
- (2) 各格子セル内での CS-LBP 値のヒストグラム (CS-LBP ヒストグラム) を算出する。
- (3) 各格子セルの CS-LBP ヒストグラムの総和が 1 になるように正規化を施す。
- (4) 格子セルごとに求められた CS-LBP ヒストグラムを連結して $M \times N \times 16$ 次元の特徴量を得る。

図 5 に格子セル数を 5×5 としたときの特徴抽出例を示す。

3.3 推定手法

本研究では、笑顔の検出器として式 (2) で表されるガウスカーネルによる非線形 SVM を用いた。

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right) \quad (2)$$

ここで、 σ はガウスカーネルのパラメータである。今、学習によって N 個のサポートベクタが得られていたとすると、カーネルトリックを用いた非線形 SVM の識別関数は、ある入力ベクトル \mathbf{x} が与えられたとき、 i 番目のサポートベクタを \mathbf{s}_i 、 i 番目のサポートベクタに対応した重み係数を α_i とすると、カーネル関数 K とバイアス項 h を用いて式 (3) のように表される。

$$y = \sum_{i=1}^N \alpha_i K(\mathbf{s}_i, \mathbf{x}) - h \quad (3)$$

ここで、笑顔検出においては、得られた出力 y をもとに、 $y \geq 0$ のときに笑顔、 $y < 0$ のときに非笑顔と判定する。次に、この笑顔検出結果から笑顔度 I_{smile} を推定することを考える。本研究では笑顔検出器の出力 y から式 (4) で示されるシグモイド関数を用いて笑顔度 I_{smile} を算出する。

$$I_{smile} = \frac{1}{1 + \exp(-\lambda \cdot y)} \quad (4)$$

ここで、ゲイン λ は笑顔度の推定感度に相当し、本研究では $\lambda = 5$ として後述の実験を実施した。

4. 笑顔検出性能の評価

本研究では、対象者が通常の音楽療法活動が可能のように非接触・非拘束な状態で撮影された映像データを用いる。そのため、対象者の顔向きが撮影環境、活動内容や姿勢に応じて変化することが想定される。また病院や福祉施設などで実際に運用されることを目指すうえでは不特定多数の対象者に対して笑顔が検出可能な汎化性能が求められる。

4.1 顔向きに対する性能評価

まずはじめに顔向きに対する笑顔の検出性能を評価することを考える。ここで評価用の画像データベースは、Web 上の公開画像や顔画像の収録を行い収集した画像群に対し、性能評価の対象を本提案手法で検出可能な顔画像に絞るために、3.1 節で述べた手法により顔検出を行い、正しく顔が検出された画像のみを抽出して構築した。ここで、本提案手法の顔検出器は、たとえば真横向き、つまり片方の目が完全に隠れてしまっているような“横顔”に対しては、顔を安定して検出することが困難であるが、3.1 節で触れたように、ほぼ真正面を向いている“正面顔”に加えて、水平方向に ± 30 度程度の“斜め顔”の顔画像を用いて顔検出器を学習したため、おおよそ学習画像と同程度の範囲内での顔向き変化に対しては顔が検出可能である。



図 6 顔向き別笑顔/非笑顔画像データベースのサンプル画像

Fig. 6 Sample images of our original smile/non-smile DB in each face direction.

このような顔検出器によって検出された顔画像群に対し，“正面顔”/“斜め顔”の分類と笑顔/非笑顔の分類を各々目視で行い，“正面顔”の笑顔/非笑顔画像データベース（笑顔：2,620 枚，非笑顔：3,620 枚）と“斜め顔”の笑顔/非笑顔画像データベース（笑顔：2,260 枚，非笑顔：4,170 枚）を各々構築した。このとき，“正面顔”と“斜め顔”との目視分類の基準は，片方の眼の目尻が顔の輪郭に達している場合は“斜め顔”と判定し，それ以外は“正面顔”とした。このようにして構築した“正面顔”の笑顔/非笑顔画像データベースおよび“斜め顔”の笑顔/非笑顔画像データベースのサンプル画像を図 6 に示す。そして，この各々の顔向きの笑顔/非笑顔画像データベースを用いて，5 分割交差確認法により顔向きごとに笑顔検出性能のオープンな評価を実施した。本評価では特徴量として，3.2 節で説明した LIH 特徴量，CS-LBP 特徴量，そしてそれらの 2 つを連結した LIH+CS-LBP 特徴量を用いて，笑顔検出性能を ROC 曲線で比較する。ここで，各々の特徴量の特徴抽出パラメータはあらかじめ実施した基礎評価実験により最も高い性能を示した以下の値とした。

- LIH 特徴量
 - 格子セル数 : 8×8
 - ヒストグラムビン数 : 4
 - 次元数 : 256
- CS-LBP 特徴量
 - 格子セル数 : 5×5
 - 近傍画素数 : 8
 - 半径 : 1
 - エンコード閾値 : 0.02
 - 次元数 : 400
- LIH+CS-LBP 特徴量
 - LIH と CS-LBP 各々のパラメータは上記に従う
 - 次元数 : 656

LIH，CS-LBP，LIH+CS-LBP 各々を特徴量とした場合の“正面顔”に対する笑顔検出 ROC 曲線を図 7 に，“斜め顔”に対する笑顔検出 ROC 曲線を図 8 に各々示す。“正面顔”の笑顔検出性能に着目すると，いずれの特徴量を用いても AUC (Area Under the Curve) > 0.97 と非常に高い検出性能を示しており，なかでも 2 つの特徴量を結合した LIH+CS-LBP 特徴量を用いたときに $AUC = 0.981039$ と最も高い性能を示した。一方，“斜め顔”の笑顔検出性能は，“正面顔”と比較すると各特徴量ともに若干検出性能

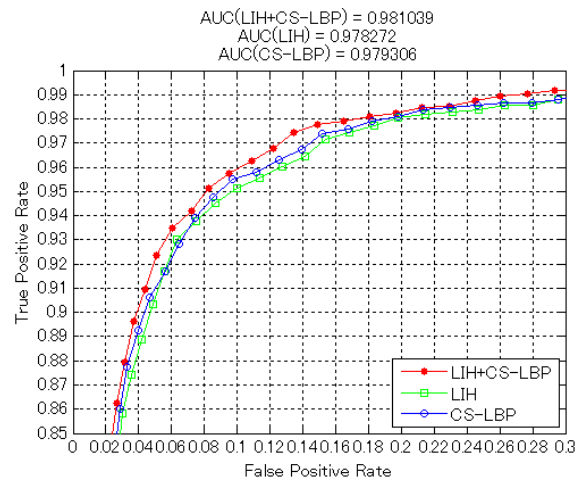


図 7 “正面顔”に対する笑顔検出 ROC 曲線の特徴量による比較
Fig. 7 Comparison of the ROC curve of each feature extraction method for frontal faces.

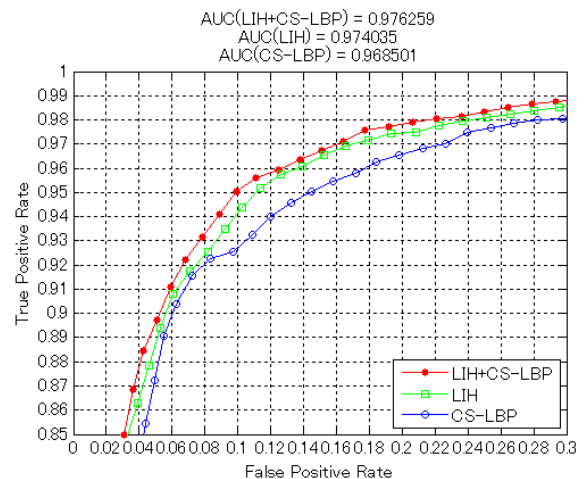


図 8 “斜め顔”に対する笑顔検出 ROC 曲線の特徴量による比較
Fig. 8 Comparison of the ROC curve of each feature extraction method for three-quarter faces.

が劣るものの，いずれの特徴量においても $AUC > 0.96$ を示し，また“正面顔”同様に LIH+CS-LBP 特徴量を用いたときに $AUC = 0.976259$ と最も高い検出性能を示した。ここで特徴量別に観察すると，CS-LBP 特徴量は他の特徴量に比べて“正面顔”と“斜め顔”の検出性能を比較した場合に劣化が大きいことから，顔向きに対しては CS-LBP 特徴量よりも LIH 特徴量の方が有効であると考えられる。これらの結果から，本研究では音楽療法効果の評価に用いる笑顔検出器を構成する特徴量としては，顔検出が可能な水平方向に ± 30 度程度以内の顔向き変化において最もロバストな笑顔検出性能を示した LIH+CS-LBP 特徴量を用いることとした。

4.2 汎化性能の評価

本研究では，提案手法が実際の臨床映像データに適用されることから，不特定人の様々な表情に対して笑顔が検

出できるような高い汎化性能が求められる。ここで、汎化性能を検証するために研究用に公開されている画像データベースを評価用画像として利用し、あわせて本提案手法と他の手法との性能を比較評価することを考える。本評価では評価用画像として、表情画像データベースとして著名な Extended Cohn-Kanade Dataset (以下、CK+データセット) [20] と、大規模な笑顔画像データベースとして知られている GENKI-4K データベース [21] を利用する。また他手法との比較としては、APIアランス・ベースの表情検出手法 [9], [10] などでも用いられ高い表情検出性能を示すことが知られており、画像認識の分野において代表的な画像特徴量の1つである LBP ベースの特徴量を笑顔検出のための特徴量として用い、本研究で採用した LIH+CS-LBP 特徴量と検出性能を ROC 曲線により比較する。ここで、LBP ベースの特徴量としては文献 [9], [10] でも提案手法との比較対象手法として用いられている Uniform LBP 特徴量 [19] を用い、特徴抽出パラメータはあらかじめ実施した基礎評価実験により最も高い性能を示した以下の値とした。

• Uniform LBP 特徴量

- 格子セル数 : 5 × 5
- 近傍画素数 : 8
- 半径 : 1
- 次元数 : 1,475

まずはじめに、多表情画像データを用いて様々な表情に対する笑顔検出器の汎化性能を評価する。CK+データセットは、表情画像データベースとして研究用途に広く利用されてきた Cohn-Kanade AU-Coded Facial Expression Database [22] をさらに拡張し、18歳から50歳の被験者123人に対する表情画像データが、無表情の状態から表情が表出されるまでの連続した画像として593シーケンス収められており、またそのうち327シーケンスについては Emotion ラベル (Angry, Contempt, Disgust, Fear, Happy, Sadness, Surprise) も同時に提供されている。本評価では、このラベル情報をもとに笑顔画像として Happy の Emotion ラベルが付与されている画像シーケンスの最終フレームの画像 (69枚) を、また非笑顔画像としては、Happy 以外の Emotion ラベルが付与されている画像シーケンスの最終フレームの画像 (258枚) と、無表情画像として各画像シーケンスの先頭フレームの画像 (327枚) をあわせて用い、その結果、評価用画像群は、笑顔画像69枚、非笑顔画像585枚の構成となった。この評価用画像群に対して、4.1節で説明した顔向きごとの笑顔/非笑顔画像データベースの全画像を学習用画像として、LIH+CS-LBP 特徴量と Uniform LBP 特徴量を各々抽出して学習した笑顔検出器を用いて笑顔の検出性能を比較評価した結果を図9に示す。いずれの特徴量を用いても $AUC > 0.99$ と非常に高い笑顔検出性能が示されており、この結果から本研究で笑顔度の定量化手法として用いる LIH+CS-LBP 特徴量を

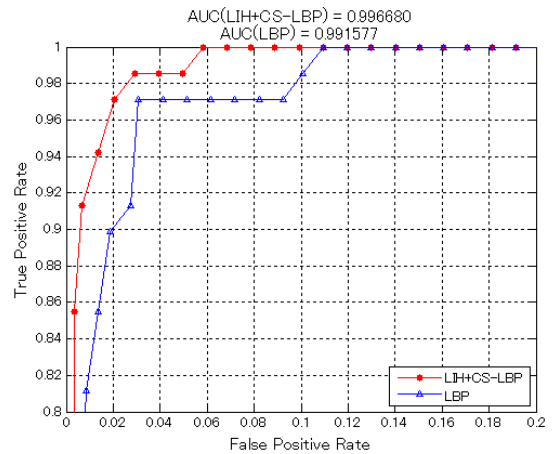


図9 CK+データセットを用いた表情に対する汎化性能比較
 Fig. 9 Comparison of generalization performance for facial expression.

用いた場合においても、従来手法と同様に笑顔以外の表情を非笑顔として高精度に検出できることを確認した。またたとえば、多方向の顔画像に対して高い表情検出性能が報告されている文献 [10] で用いられている LGBP 特徴量と比較すると、16サブブロックのLGBP特徴量を用いた場合に $AUC = 0.99556$ と提案手法と同等の検出性能を示すことを確認したが、このとき、特徴量の次元数は37,760次元と656次元のLIH+CS-LBP特徴量に比べ60倍弱であり、臨床応用を考えた場合に重要な要素の1つである笑顔検出のリアルタイム性の確保という点においては課題があると考えられる。

次に、不特定人に対する笑顔検出器の汎化性能を評価する。GENKI-4K データベースは、研究用に公開されている笑顔/非笑顔画像データベースで、CK+データセットのように実験室などの整えられた撮影環境下で収集された画像ではなく、Webベースで画像が収集されており多人種、多撮像環境、多解像度の画像かつ自然な表情でデータベースが構成されているため、笑顔検出器の汎化性能を評価するのに適している。このGENKI-4Kデータベースの全画像データ4,000枚のうち、本研究の顔検出器で顔が検出できた3,653枚 (笑顔: 2,010枚, 非笑顔: 1,643枚) を、4.1節で説明した独自画像データベースに追加し、より大規模な笑顔/非笑顔画像データベース (笑顔: 6,890枚, 非笑顔: 9,433枚) を構築した。そして、この大規模笑顔/非笑顔画像データベースを対象として、LIH+CS-LBP 特徴量、Uniform LBP 特徴量各々について5分割交差確認法でオープンな比較評価を実施した結果を図10に示す。このように大規模な画像データを用いた場合でも、LIH+CS-LBP 特徴量を用いた場合に $AUC = 0.975635$ 、Uniform LBP 特徴量を用いた場合に $AUC = 0.97058$ と両特徴量ともに高い笑顔検出性能を示し、不特定人に対してもロバストに笑顔が検出可能であることを確認した。よって本節の2つの評価実験を通じ、LIH+CS-LBP 特徴量は Uniform LBP 特

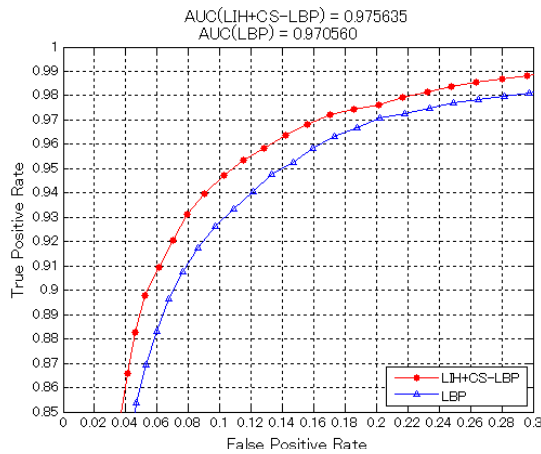


図 10 大規模笑顔/非笑顔データベースを用いた不特定多数人に対する汎化性能比較

Fig. 10 Comparison of generalization performance for person independence.

微量の半分以下の次元数で、顔表情および不特定人に対して同等の汎化性能を示すことを確認した。

4.3 笑顔度推定評価

次に、FEED データベース [23] を用いて笑顔度を推定した結果を示す。FEED データベースは、笑顔、怒りや恐れなど Ekman ら [3] が定義した 6 つの表情ラベルが付与されており、さらにこれらの表情画像が 18 人のサブジェクトに対して無表情から特定の表情への画像シーケンスとして、各人につき 3 試行分提供されている。また、画像シーケンスが 100 から 150 フレーム程度と比較的長いフレーム数で構成されているため、表情の微妙な変化度合いを評価する目的に非常に適している。図 11、図 12 に提案した笑顔度推定手法を用いて FEED データベースに含まれる、2 人の笑顔画像シーケンス（サブジェクト番号 0006 の第 1 試行、サブジェクト番号 0018 の第 3 試行）に対して笑顔度を推定した結果を各々示す。両者ともに無表情から笑顔への表情変化が笑顔度として非常によく数値化されており、特に図中赤枠で示した付近の微妙な表情変化もよく推定できていることが分かる。また、図 11 に示したサブジェクト番号 006 の第 1 試行の画像シーケンスは、無表情から笑顔へ、笑顔から無表情へ、そして最後にまた無表情から笑顔へと表情が変化しているが、本提案手法により各表情変化を追従できていることが確認できる。また提案手法を例で示した 2 人の他の試行、およびその他 16 人の全試行についても同様に適用した結果、実際には笑顔の表情を示していなかったサブジェクト番号 0017 の第 1 試行、サブジェクト番号 0011 の第 2 試行、サブジェクト番号 0014 の第 3 試行の画像シーケンス以外は、図 11、図 12 に示したのと同様に対象者の笑顔の変化度合いに応じて笑顔度が推定できていることを確認した。

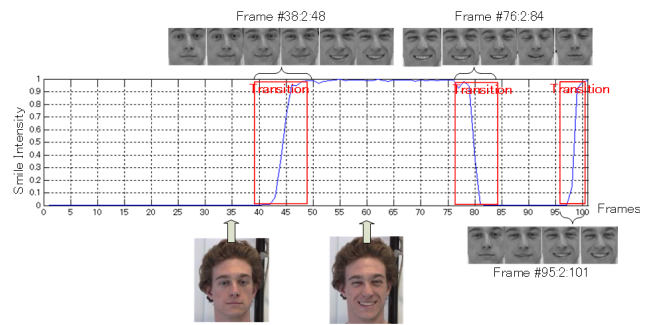


図 11 サブジェクト番号 006, 第 1 試行の画像シーケンスに対する笑顔度推定結果

Fig. 11 Result of the smile intensity estimation for images of Subject #0006, Trial #1.

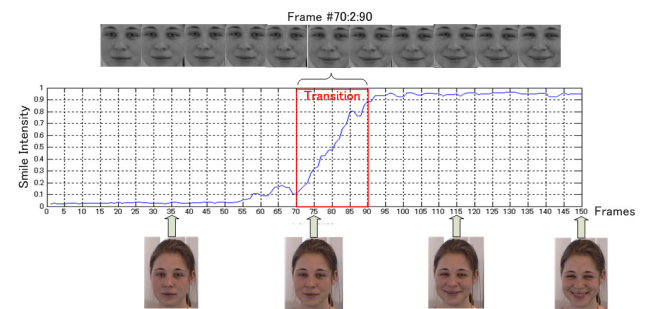


図 12 サブジェクト番号 0018, 第 3 試行の画像シーケンスに対する笑顔度推定結果

Fig. 12 Result of the smile intensity estimation for images of Subject #0018, Trial #3.

5. 療法効果評価への応用

次に、提案した笑顔度推定手法を実際の臨床映像に適用し、音楽療法効果の評価へ応用することを考える。本研究は茨城県立医療大学倫理委員会の承認を得て実施され、映像を研究利用するための患者へのインフォームド・コンセントは以下の 2 段階で行われた。1 段階目のインフォームド・コンセントは、茨城県立医療大学附属病院で実施される音楽療法を受けるすべての患者に対して初回導入時に適用されるものである。ここでは、映像記録を行うこと、その映像データを参加者のアセスメントや療法内容の向上のために用いること、非公開で院外には持ち出さないこと、希望者は背面位置や撮影視野外に座席を移動して顔を映さないことも可能であることなどを説明して承諾を得る。それに加えて、2012 年 1 月からは本研究への協力の呼びかけを開始し、それに応じた参加者もしくはその家族に対して個別に 2 段階目のインフォームド・コンセントを実施した。ここでは、研究の目的と目的外で映像データを使用しないこと、映像データから抽出された匿名化数値（本研究においては、笑顔度など）が公表される場合があることなどを口頭・書面で説明して承諾を得た。結果、同年 8 月の時点で 3 人の研究協力（つまり、2 段階目のインフォームド・コンセントへの承諾）が得られた。この 3 人の映像データ

を視聴し、退院までの期間が短かった患者1人、座位姿勢が困難な症状で着座中は顔が大きく俯いてしまい、映像データからはほとんど顔および表情が観測できなかったため、本研究の顔検出器の現状性能では検出が困難であった1人を各々対象から除き、残りの1人について機器の不具合で記録できなかった回や着座位置の関係で顔が確認できなかった回などを除外した結果、比較的長期にわたる6回のセッションで良好な映像が確認できた。そこで本研究では、この1人を対象に笑顔度の推定を試みることにした。ここで解析対象となった患者は、NMDA 受容体関連脳炎を発症した10代女性で、初回参加の3月時点では覚醒状態が安定せず、四肢麻痺があり、日常生活動作に全介助が必要な状態であった。そこで、覚醒水準の向上とコミュニケーション反応を促す目的で担当作業療法士の要請を受けて音楽療法が導入され、月に隔週2回の頻度で3月から8月までに実施された12回の音楽療法セッションのうち10回に参加した。しかし前述のとおり、撮影機器の不具合や着座位置の影響により、実際に研究利用可能な映像データが得られたセッションは、3、4、5、7月の各1回および8月の2回の計6回であった。

5.1 臨床映像からの笑顔度推定

本研究が対象とした音楽療法セッションは各回の参加患者が4~7人、音楽療法士、スタッフらが円陣に着座した状態で行われ、患者の着座位置は特に定めなかった。そのため、患者が死角に入るのを極力防ぐため、約50°異なる方向から円陣を望むように2台のビデオカメラ（Sony製HDR-CX560）を設置し、プログレッシブ・ハイビジョン画質（1,920×1,080 pixel）で撮影した。円陣の直径は約3m、各カメラは円陣中心から距離約4.5m、高さ約1.2mの位置に設置し、円陣全体が撮影されるように画角を調整し撮影した。また、撮影を患者になるべく意識させない配慮として、患者の着座位置を特に定めないことに加えて、三脚などの撮影機材を用いず、ピアノや音響機器などの室内備品の上に無造作にカメラを設置した。

1回の音楽療法セッションは、各月ごとに季節にちなんだ唱歌を取り上げる“季節の歌”、簡単に演奏できる打楽器を音楽に合わせて打ち鳴らす“楽器演奏”や各人が好みの曲をマイクを持って歌う“カラオケ”などいくつかのプログラムから構成され1時間程度実施される。プログラム内容によっては対象者の顔角度が頻繁に変わったり、他の患者の影に入ったりするために安定して顔検出できない場合があった。今回はその中で最も顔検出が安定して処理できた“季節の歌”プログラムを笑顔度推定の対象プログラムとした。図13に対象とした患者の全6回の音楽療法セッションのうちある2日の“季節の歌”プログラムの映像データに対して、笑顔度推定を行った結果を例として示す。本提案手法では特に患者を拘束していないため、従来同様に自

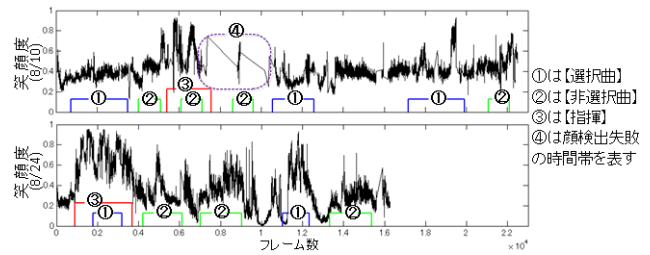


図13 “季節の歌”プログラム中の笑顔度推定結果例

Fig. 13 Smile intensity during the “seasonal song” activity.

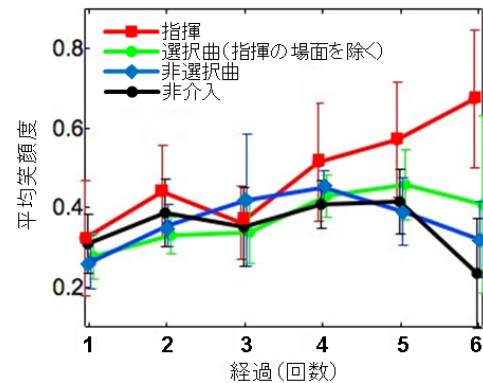


図14 介入内容ごとの平均笑顔度の推移

Fig. 14 Transition of the average smile according to each intervention.

然な形で音楽療法セッションが行える一方で、顔向きの変化などに起因して顔検出に失敗した区間（図13中4の区間）もあるものの、ほぼ全域にわたって笑顔度が推定できていることが分かる。ここで“季節の歌”プログラムでは、季節にちなんだ唱歌2曲を取り上げ、音楽療法士は患者に、どちらかの曲を選んで皆を指揮しながら歌うように働きかける構成になっている。通常患者は自分が好きな方の曲を選択して、ピアノ伴奏に合わせて歌唱の指揮を行う。このとき音楽療法士やスタッフは障害の状況に応じて患者の指揮を介助する。この介入をすべての患者で順番に行うので、1人の患者にとっては、自分が好きな曲を選んで指揮して歌う場面【指揮】（図13中3の区間）、自分が選んだのと同じ曲を指揮せず歌う場面【選択曲】（図13中1の区間）、自分が選んだのとは違う曲を歌う場面【非選択曲】（図13中2の区間）、および、それら以外の積極的介入がない場面【非介入】という4種類の区間が生じることになる。本研究では、音楽療法の経過とともに各々の介入区間ごとの平均笑顔度がどのように変化したかを評価することを考える。図14に介入内容ごとの平均笑顔度の音楽療法経過回数による推移を示す。

5.2 統計的検定による推定結果の評価

まず最初に、大まかなデータの傾向を知るために、経過回数（1~6）と介入内容（選択曲、非選択曲、指揮、非介入）を要因とする二元配置分散分析を行った。その結果（表1

表 1 二元配置分散分析結果
Table 1 Result of two-way ANOVA.

変動要因	偏差平方和	自由度	平均平方	F 値	P 値	F(0.95)
全変動	11.273	651				
行間変動 (介入内容)	1.299	3	0.433	41.898	1.086E-24	2.619
列間変動 (経過回数)	1.570	5	0.314	30.380	1.062E-27	2.228
交互作用	1.914	15	0.128	12.347	3.997E-27	1.682
誤差変動	1.490	628	0.010			

表 2 Tukey-Kramer 法による多重比較結果 (** $p < 0.01$, * $p < 0.5$)
Table 2 Result of Tukey-Kramer's test. (** $p < 0.01$, * $p < 0.5$)

	#1 指揮	#2 指揮	#3 指揮	#4 指揮	#5 指揮	#6 指揮
#1 指揮						
#2 指揮	*					
#3 指揮						
#4 指揮	**		*			
#5 指揮	**	*	**			
#6 指揮	**	**	**	**		

	#1 選択曲	#2 選択曲	#3 選択曲	#4 選択曲	#5 選択曲	#6 選択曲
#1 選択曲						
#2 選択曲						
#3 選択曲						
#4 選択曲						
#5 選択曲	*	**	*			
#6 選択曲						

	#1 非選択曲	#2 非選択曲	#3 非選択曲	#4 非選択曲	#5 非選択曲	#6 非選択曲
#1 非選択曲						
#2 非選択曲						
#3 非選択曲	**					
#4 非選択曲	**					
#5 非選択曲	**					
#6 非選択曲				*		

	#1 非介入	#2 非介入	#3 非介入	#4 非介入	#5 非介入	#6 非介入
#1 非介入						
#2 非介入						
#3 非介入						
#4 非介入	**					
#5 非介入	**		*			
#6 非介入		**	**	**	**	

参照), $p < 0.01$ の有意水準で音楽療法の経過回数による平均笑顔度の違いおよび介入内容による平均笑顔度の違いが認められたが, 同時にそれらの交互作用も認められた。

そこでさらに詳細な分析を行うため, Tukey-Kramer 法による多重比較検定を実施した。するとたとえば, 表 2 最上段に示すように, 【指揮】時の平均笑顔度における回数による漸増は有意であることが分かったが, それと同時に表 2 最下段に示すように, 【非介入】の平均笑顔度も有意に変動していることが分かった。

そこで, 各回での【非介入】時の平均笑顔度をその回の対象者の表情(または感情)ベースラインと考え, 他の介入内容の平均笑顔度からその値を減ずることで平均笑顔度を正規化し, その正規化平均笑顔度に対して Dunnett 法による多重比較検定を実施した。その結果, 図 15 に示すように【指揮】時は後半 3 回で継続して $p < 0.01$ の有意水準で正規化平均笑顔度の増加が顕著に認められ, 最終回に至っては, 【選択曲】【非選択曲】においても各々 $p < 0.01$, $p < 0.05$ の有意水準で笑顔が増えたことが認められた。

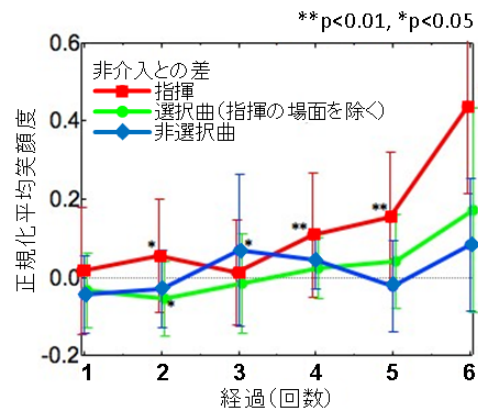


図 15 Dunnett 法による多重比較結果
Fig. 15 Result of the Dunnett's test.

表 3 主観評価と平均笑顔度との相関
Table 3 Correlation to the subjective assessment.

主観評価項目	指揮	選択曲	非選択曲	非介入
積極性	0.801	0.883	0.353	0.025
持続性	0.609	0.794	0.193	-0.059
協調性	0.906	0.963	0.141	-0.089
情動表出	0.868	0.942	0.156	-0.041
状況適切性	0.971	0.975	0.221	-0.099
知的機能	0.943	0.981	0.169	-0.158
発声発語	-0.083	0.114	-0.784	-0.403
歌唱	0.096	0.202	0.605	0.470

5.3 主観評価との比較

最後に, 従来の一般的な評価方法と笑顔度推定結果との関連性について調査した結果について述べる。従来の評価では, 音楽療法中の患者の様子を記憶や記録映像を確認しながら, あらかじめ用意された主観評価表に記載の各評価項目に関して段階評価するなどの方法がとられてきた。ここで主観評価表は音楽療法セッションごとにその音楽療法に参加した複数の音楽療法士によって評価されるため, 各評価項目における平均値を主観評価値として平均笑顔度との相関を調査した。表 3 に全 6 回の音楽療法における, 各評価項目の主観評価値と提案手法により推定された平均笑顔度との相関係数を示す。評価項目の中でも, 特に情動に関する情動表出, 状況適切性や協調性, 積極性, 知的機能などと平均笑顔度との相関が非常に高いことが分かる。また, ここで介入の内容に着目して観察してみると【指揮】や【選択曲】など介入レベルが高いほど相関が高くなっていることから, 音楽療法士の積極的な介入により, 患者の社会性やコミュニケーション力が改善された結果, 表情変化が豊かになっていったとも考えられる。一般的に考えても笑顔は, 情動反応や他者とのコミュニケーション場面で表出されることが多いと想定されるが, この結果からもその妥当性が示され, 従来の評価方法との比較によっても笑顔度と患者の療法効果による症状の回復度合い(特に情動に関わる)との高い関連性が示されたと考える。

6. まとめ

本研究では、SVMにより映像情報から笑顔度を推定する技術を開発し、その技術を用いて音楽療法の記録映像のみから患者の笑顔度を推定し定量化することで、その時間的・量的推移から音楽療法の効果を客観的に評価する方法について検討した。

笑顔度推定に関しては、顔器官を検出することなく顔の_Apiランス情報のみからSVMにより笑顔の検出と笑顔度を推定することを目指し、顔向きに対する基本的な笑顔の検出性能を独自に構築した顔向き別の笑顔/非笑顔画像データベースを用いて検証した結果、LIH+CS-LBP特徴量を用いた場合に、“正面顔”において $AUC = 0.981039$ ，“斜め顔”において $AUC = 0.976259$ といずれの顔向きにおいても高い笑顔検出性能を達成した。

また汎化性能の評価においては、目的別に2種類の研究用公開データベースを用い、他手法の比較とあわせて評価した結果、CK+データセットを対象とした場合に $AUC = 0.99668$ ，GENKI-4K データベースを対象とした場合に $AUC = 0.975635$ となり、LIH+CS-LBP特徴量を用いると従来手法の半分以下の次元数で同等の笑顔検出性能を示すことを確認し、またFEEDデータベースを用いた笑顔度推定の評価では、人物の表情変化画像シーケンスを用いて本手法が無表情から笑顔への微妙な表情変化を笑顔度として定量化可能なことを示した。

次に、開発した笑顔度推定手法を用いた音楽療法効果の評価実験では、半年間計6回にわたって継続的に撮影された実際の臨床現場の音楽療法記録映像を用いて、対象患者の顔向きが正面付近の場合には、顔検出および笑顔度の推定が安定して行えることを確認し、またその測定結果を多重比較検定することで、音楽療法の経過回数にともない患者の平均笑顔度が有意に増加していることを示した。また、これと並行して実施した従来の音楽療法士による主観評価結果と本提案手法による笑顔度の定量化結果が、特に情緒を示す評価項目において高い相関があることが確認された。ただ一方で、実際の臨床映像を対象とした場合、現状の顔検出および笑顔度推定技術では多種多様な症状をかかえた患者に対し実用上十分に安定して笑顔度を数値化できるほどの性能を達成したとはいえ、今後アルゴリズムのさらなる改良、撮像方法の工夫などが必要であることも確認した。

本研究は、実際に音楽療法の現場での患者の情動賦活を笑顔度という指標を介して客観的に定量化した初めての報告である。今後需要の増大が予想される音楽療法を含む医療福祉サービス一般について、その効果をこのように客観評価することが可能になれば、より合理的に医療福祉に携わる多職種間での意思疎通、情報交換、スキル向上を図ることが可能になると期待できる。この研究がその試みの1

つとして、将来の医療福祉サービスの質の維持と向上のための一助となれば、幸いである。

7. 今後の展望

本稿では継続してデータが蓄積でき、笑顔度変化の経過を解析できた患者数は1人であった。そのため、療法効果に関する今回の結果が多くの患者に共通するかについては、現状では多くを語ることはできず、臨床現場の諸制約のもと粘り強くデータを蓄積して検証を重ねていくことが今後の大きな課題の1つである。その検証過程では、非接触・非拘束の状態、より安定して顔検出・笑顔度推定ができるように種々の要因（撮影・照明方法、画像解像度、ターゲットとするプログラムなど）を体系的に検討し、最適化を図っていくことも必要である。また、2.3節で述べたように、評価尺度として笑顔だけでなく、怒りや悲しみなどの不快感情を利用するための表情検出手法の検討および不快感情を評価尺度として用いる場合の研究デザインの構築について検討し、新たな療法効果の客観的評価手法についても検討していきたい。また、本研究では脳機能障害にともない精神機能が当初低調であった症例を取り上げたが、対象を脳卒中や認知症の患者へも広げていくことを考えてみると、たとえば脳卒中の後遺症として見られる感情失禁のような病理解例に対しては本提案手法を単純にそのまま適用することはできない。従来、感情失禁はその有/無や笑い/泣きの主観的鑑別に基づいて神経心理学的な手法で研究されてきた[24]が、本提案手法の適用を同一刺激下での感情失禁患者と健常者（たとえば、患者とともに音楽療法へ参加した家族）へと拡大し、各々の表情表出の差異比較へと発展させ表情変化の共起性を考慮することができれば、感情失禁の鑑別やその程度をより客観的なものにするのができ、それによって病態機序のより詳細な解明への道が拓ける可能性があると考えている。また、認知症ではその進行とともに感情が平板化して表情表出が乏しくなるといわれている。その進行の程度を評価する目的で、古くはPhiladelphia Geriatric Center Affect Rating Scale[25]や近年ではPsychological Assessment Scale by Facial Expression for Demented People[26]などの評定スケールが提案されているが、いずれの手法も評定者の主観に基づく採点を基礎においており、その信頼性は評定者の熟練度などに大きく依存する。本研究で示した笑顔検出技術を改良し、多様な表情で確立できれば、認知症の進行にともなう表情の平板化を何年にもわたって客観的に評価できる可能性がある。たとえば、実施プロトコルを適切に設定し、仮に施設利用者を定点観測した映像データを得ることが可能となった場合、そこで得られた映像データは従来と異なって、評定者の熟練度などに依存しないため複数の施設間で相互に比較できる。大規模疫学的調査などでこうした手法を導入することは、効率性・客観性の観点から今

後重要性を増していくものと考えられる。

謝辞 本研究を進めるにあたり、研究協力にご同意いただいた茨城県立医療大学付属病院の音楽療法参加者の皆様、および研究の場を与えてくださった病院関係者の皆様に深く感謝いたします。

参考文献

- [1] 総務省統計局：人口推計（平成23年10月1日現在），総務省統計局（オンライン），入手先〈<http://www.stat.go.jp/data/jinsui/2011np/index.htm>〉（参照2012-12-01）。
- [2] Japan Music Therapy Association：音楽療法とは，Japan Music Therapy Association（オンライン），入手先〈<http://www.jmta.jp>〉（参照2012-03-22）。
- [3] Ekman, P. and Friesen, W.V.: *Unmasking the face. A guide to recognizing emotions from facial cues*, Prentice Hall (1975).
- [4] Fasel, B. and Luetttin, J.: Automatic Facial Expression Analysis: A Survey, *Pattern Recognition*, Vol.36, No.1, pp.259–275 (1999).
- [5] Yonemura, K. and Sugiura, A.: An assessment of cerebral disease by using face image synthesis, *Engineering in Medicine and Biology Society, 2000, Proc. 22nd Annual International Conference of the IEEE*, Vol.1, pp.406–414 (2000).
- [6] Takasu, S., Sugiura, A. and Yonemura, K.: Suggestion of the dependency diagnosis support system focused on the expression change, *World Congress on Medical Physics and Biomedical Engineering 2006, IFMBE Proc.*, Magjarevic, R. and Nagel, J. (Eds.), Vol.14, pp.400–404, Springer Berlin Heidelberg (2007).
- [7] 二俣 泉：音楽療法の設計，春秋社 (1995)。
- [8] Deniz, O., Castrillon, M., Lorenzo, J., Anton, L. and Bueno, G.: Smile Detection for User Interfaces, *ISVC'08: Proc. 4th International Symposium on Advances in Visual Computing, Part II*, pp.602–611, Springer-Verlag, Berlin, Heidelberg (2008).
- [9] Shan, C., Gong, S. and McOwan, P.W.: Facial expression recognition based on local binary patterns: A comprehensive study, *Image and Vision Computing*, Vol.27, No.6, pp.803–816 (2009).
- [10] Moore, S. and Bowden, R.: Local binary patterns for multi-view facial expression recognition, *Computer Vision and Image Understanding*, Vol.115, No.4, pp.541–558 (2011).
- [11] Whitehill, J., Littlewort, G., Fasel, I., Bartlett, M. and Movellan, J.: Toward Practical Smile Detection, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.31, No.11, pp.2106–2111 (2009).
- [12] Viola, P. and Jones, M.: Rapid Object Detection using a Boosted Cascade of Simple Features, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol.1, pp.511–518 (2001).
- [13] Cortes, C. and Vapnik, V.: Support-vector networks, *Machine Learning*, Vol.20, No.3, pp.273–297 (1995).
- [14] 前田英作：痛快!サポートベクトルマシン—古くて新しいパターン認識手法，*情報処理*，Vol.42, No.7, pp.676–683 (2001)。
- [15] Shimada, K., Noguchi, Y., Sasahara, H., Yamamoto, M. and Tamegai, H.: Detection of Driver's Face Orientation for Safety Driving Assistance, *Transactions of Society of Automotive Engineers of Japan*, Vol.41, No.3, pp.775–780 (2010).
- [16] Scholkopf, B., Burges, C.J.C. and Smola, A.J.: *Advances in Kernel Methods*, The MIT Press (1998).
- [17] Matsukawa, T., Hidaka, A. and Kurita, T.: Automatic Evaluation of Degree of Spectators' Satisfaction in Video Sequences Based on Their Facial Expression and Face Directions, *IPSJ*, Vol.50, No.12, pp.3222–3232 (2009).
- [18] Heikkilä, M., Pietikäinen, M. and Schmid, C.: Description of Interest Regions with Center-Symmetric Local Binary Patterns, *Computer Vision, Graphics and Image Processing*, Kalra, P. and Peleg, S. (Eds.), Lecture Notes in Computer Science, Vol.4338, pp.58–69, Springer Berlin Heidelberg (2006).
- [19] Ojala, T., Pietikainen, M. and Maenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.24, No.7, pp.971–987 (2002).
- [20] Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z. and Matthews, I.: The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression, *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp.94–101 (2010).
- [21] Machine Perception Laboratory: The MPLab GENKI Database, GENKI-4K Subset, University of California, San Diego (online), available from 〈<http://mplab.ucsd.edu>〉 (accessed 2010-06-22).
- [22] Kanade, T., Cohn, J.F. and Tian, Y.: Comprehensive database for facial expression analysis, *Proc. 4th IEEE International Conference on Automatic Face and Gesture Recognition*, pp.46–53 (2000).
- [23] Wallhoff, F.: Facial Expressions and Emotion Database, Technische Universität München (online), available from 〈<http://www.mmk.ei.tum.de/~waf/fgnet/feedtum.html>〉 (accessed 2010-06-22).
- [24] Lauterbach, E.C., Cummings, J.L. and Kuppaswamy, P.S.: Toward a more precise, clinically-informed pathophysiology of pathological laughing and crying, *Neuroscience & Biobehavioral Reviews*, Vol.37, No.8, pp.1893–1916 (2013).
- [25] Lawton, M.P.: Quality of life in Alzheimer disease, *Alzheimer Disease & Associated Disorders*, Vol.8, No.3, pp.138–150 (1994).
- [26] 矢野啓明, 高橋伸佳, 斯波純子ほか：重度認知症患者における視線と表情による簡易心理評価スケールの開発，高次脳機能研究（旧失語症研究），Vol.32, No.2, pp.312–319 (2012)。



嶋田 敬士

1995年山口大学工学部知能情報システム工学科卒業。1997年奈良先端科学技術大学院大学情報科学研究科修士課程修了。同年旭化成工業株式会社（現、旭化成株式会社）入社。2009～2012年産業技術総合研究所招聘研究員。現在、旭化成株式会社融合ソリューション研究所チーフエンジニア。2010年より筑波大学大学院システム情報工学研究科博士課程に在籍。映像情報メディア学会会員。



福井 和広 （正会員）

1986年九州大学工学部機械工学科卒業。1988年同大学大学院工学研究科修了。同年（株）東芝入社。同社研究開発センター主任研究員を経て、現在、筑波大学大学院システム情報工学研究科教授。パターン認識・コンピュータビジョンの理論と応用に関する研究に従事。工学博士。2002年電子情報通信学会論文賞、2006年電子情報通信学会ISS活動功労賞受賞。電子情報通信学会、IEEE各会員。

山田 亨

1991年筑波大学大学院物理学研究科博士課程修了、理学博士。1992年通産省工業技術院電子技術総合研究所研究員。2001年独立行政法人産業技術総合研究所脳神経情報研究部門主任研究員。2010年より同ヒューマンライフテクノロジー研究部門主任研究員。近赤外脳機能計測の計測信頼性向上技術の開発研究に従事する傍ら病院ボランティアとして音楽療法に参画。日本神経科学会、日本生体医工学会、日本ヒト脳機能マッピング学会等各会員。

高崎 友香

2009年茨城県立医療大学卒業。2011年筑波大学大学院人間総合科学研究科博士前期課程修了。同年茨城県立医療大学付属病院入職。作業療法士。日本作業療法士協会会員。



野口 祥宏 （正会員）

1985年神戸大学工学部計測工学科卒業。同年旭化成入社。1993～1995年カリフォルニア大学バークレー校客員研究員。2009～2012年産業技術総合研究所招聘研究員。現在、旭化成融合ソリューション研究所グループ長。IEEE、電子情報通信学会、電気学会各会員。博士（工学）。

山崎 郁子

国立療養所東京病院附属リハビリテーション学院卒業。桐朋学園大学音楽学部演奏学科声楽専攻卒業。テキサス州立女子大学大学院表現芸術学研究科修了。茨城県立医療大学助教授、帝京平成大学教授、神戸大学教授、東京工科大学教授を経て、現在、国際医療福祉大学大学院特任教授。芸術修士。作業療法士。認定音楽療法士。