

弱視者の笑顔認識の傾向に関する調査

中山 雄貴^{1,a)} 松原 正樹² 鈴木 健嗣³ 寺澤 洋子^{2,b)}

概要: 本研究では、笑顔可聴化システムを用いて弱視者による他者の笑顔認識を支援することを目指す。可聴化とは、本来音ではない情報を非言語音に変換して情報伝達を行う方法である。笑顔可聴化システムは入力データとして顔面の表面筋電位信号を用いており、笑顔時の表面筋電位信号をリアルタイムに可聴化する。弱視者による他者の笑顔認識能力を調べる調査を通して、弱視者は音声を伴わず歯が見えていないような度合いの小さい笑顔や正面以外の角度で表出される笑顔の認識に困難があることが示唆された。

YUKI NAKAYAMA^{1,a)} MASAKI MATSUBARA² KENJI SUZUKI³ HIROKO TERASAWA^{2,b)}

1. はじめに

他者の表情を正しく認識することは、円滑なコミュニケーションにおいて重要である。社会生活を営む上で、他者とのコミュニケーションは欠かすことのできない要素である。表情は重要なノンバーバルコミュニケーションの一種である。表情は感情を表しており、感情に伴う表情表出は文化や国籍に関わらず全人類で共通する普遍的なものであるとされている。表情に普遍性があることにより、表情から感情を推測することが可能である。相手の表情を正しく認識することが、円滑なコミュニケーションにおいて重要である。

しかし、弱視者のように他者の表情を認識することに困難を持つ人もいる。厚生労働省の調べによれば日本には20万人以上の弱視者がおり視覚障害者の6割以上を占めている[1]。弱視の代表的な症状としては矯正しても視力が低い低視力、周辺視野や中心視野の欠損といった視野障害が挙げられる[2]。彼らは視覚情報から他者の表情を認識することが難しく、声をはじめとする聴覚情報を元に他者の表情を推測している。そのため、他者が声を発していない状況においては他者の表情を認識することが難しく、対面コミュニケーションに困難が生じている[3]。

そこで本研究では、笑顔可聴化システムを用いて弱視者による音声を伴わない笑顔の認識の支援を目指す。可聴

化とは「コミュニケーションや解釈を容易にするための、データの関係性を音響信号により知覚される関係性に変換することである」と定義されている[4]。この定義に基づき、可聴化は本来音ではない情報を非言語音に変換して伝達を行う。本笑顔可聴化システムにおいては、顔面の筋電位を入力データとし、笑顔時に音を出力する。

笑顔可聴化システムの有効性検証にさきかけ、視覚情報のみを用いた弱視者の笑顔認識能力について静止画と動画を用いて調査を行った。本稿では、弱視者の笑顔認識能力に関する調査および笑顔可聴化システムの構成について述べる。

2. 関連研究

2.1 感情と表情

笑顔をはじめとする表情は、私達人間が持つ感情をよく表すものである。この感情についての考え方として、カテゴリー説と次元説という2つの重要な考え方が存在する。

カテゴリー説はEkmanらによって提唱されており、人の表情には「喜び、怒り、悲しみ、恐怖、嫌悪、驚き」という6つの基本的表情が存在し、これらは国や文化に関わらず表出、認識されると述べている[5]。カテゴリー説に当てはめると、笑顔は一般的に喜びの感情を表している表情であると考えられる。

一方で次元説は、横軸を「快・不快」、縦軸を「覚醒・睡眠」とした座標軸上で感情を連続的な変化として捉える。例えば、カテゴリーとしては同じ「喜び」であってもそこには「わくわくする喜び」「ほっとした喜び」のように細かな違いが存在しており、次元説では座標軸を用いること

¹ 筑波大学図書館情報メディア研究科
Tsukuba, Ibaraki 305-0006, Japan

² 筑波大学図書館情報メディア系

³ 筑波大学システム情報系

a) yuki@slis.tsukuba.ac.jp

b) hiroko@slis.tsukuba.ac.jp

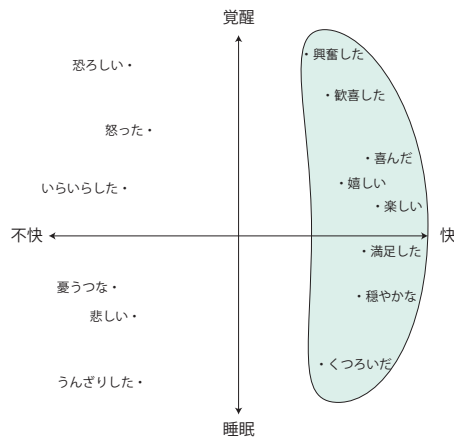


図 1: 感情の円環モデル (色付き部分は笑顔に関係すると考えられる感情の範囲)

で同じカテゴリーの感情の連続的な違いを表すことが出来る。Russellはこの座標軸を元に感情がどの位置に来るのかを形容詞で示すモデルを作成した。これを感情の円環モデルという [6]。円環モデルを図 1 に示す。笑顔は円環モデル上において快方向の広い範囲の感情に対応する表情であると考えられる。(図 1 色付き部分)。そこで本研究においては、表情の中でも笑顔に着目し、可聴化を行うこととした。

2.2 表情の可聴化と表面筋電位

本笑顔可聴化システムでは、表情の動きを捉える特徴量として筋電位信号を用いている。筋電位信号とは筋肉の活動の過程で筋繊維を収縮させるために発生する活動電位のことである。筋電位信号はその筋肉の活動の度合いによって振幅が変化し、より活発に活動するほど振幅が大きくなる。

筋電位を計測する方法は用いる電極によって大きく 2 つに分かれる [7]。1 つは皮膚表面に電極を取り付けて計測する方法で、この方法で計測される筋電位信号を表面筋電位信号と呼ぶ。もう 1 つは皮膚から筋内に針電極を刺入する方法で、筋内の電位変化を高い空間分解能で識別できるが、人体に対して侵襲性が高い。本研究では、人体に対して侵襲することなく測定出来る表面筋電位信号を用いる。

本研究において表面筋電位を用いる利点として、空間的な制限がないこと、リアルタイム性が高いことという 2 点が挙げられる。

過去の表情可聴化研究において、Patil ら [8] や Funk ら [9] は、表情の可聴化に画像処理を用いている。彼らは顔をいくつかの部位にわけ、それぞれの部位のオプティカルフローを用いて表情の可聴化を行った。

しかし、このような画像処理を用いた可聴化にはカメラが必要であり、カメラと顔の位置関係による空間的な制限が大きくなる。また表情を認識するためには顔がカメラと

正対する必要があり首をまわせないといった姿勢の制約や、カメラの存在によって生じる視界の制限もある。

一方表面筋電位信号を計測する電極は小型で軽量かつ視界に入らない位置に装着可能なため、そのような空間、姿勢、視界の制約がほとんどない。また、筋活動が表情の変化を起こし、それを他者が視覚で認知するという順序を考えれば、筋電位信号を用いると視覚情報を用いるよりも早く表情の変化を捉えられることがわかる。よって、リアルタイム性も高いと考えられる。

筋電位信号を用いた笑顔の識別は既に行われており [10]、本研究でもこの手法を元にして笑顔の可聴化を行う。

3. システム概要

本笑顔可聴化システムでは、左右の前頭部および側頭部にて計測した筋電位信号を入力として、リアルタイムに可聴化音を出力する。本システムのフロー図を図 2 に示す。

本システムは信号処理部、表情識別部、可聴化部の 3 つのモジュールから構成される。信号処理部 (Signal Processing Module) ではノイズ除去のためのフィルタ処理および特徴量の抽出を行う。表情識別部 (Facial Expression Recognition Module) では表情識別およびそのための学習を行う。可聴化部 (Sonification Module) では表情識別結果と特徴量を用いて可聴化を行うのである。可聴化音は表情識別部で笑顔と識別された場合にのみ再生される。

信号処理部および表情識別部は C#, 可聴化部はリアルタイムでの音合成に優れる SuperCollider を用いて実装した。信号処理部および表情識別部から可聴化部へのデータの受け渡しは OSC (Open Sound Control) 通信 [11] を用いた。

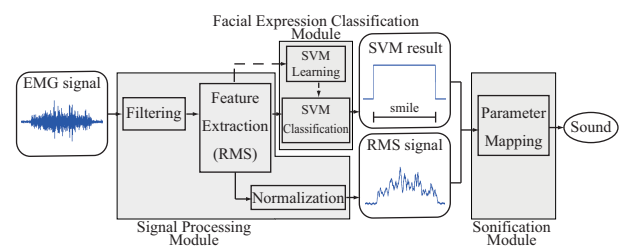


図 2: システムフロー図

3.1 信号処理部

信号処理部では計測した筋電位信号に対しノイズ除去のためのフィルタ処理を施した後、特徴量の抽出を行う。特徴量は信号の時間的な振幅の変化を表す RMS (Root Mean Square) を用いる [12]。

フィルタ処理はコムフィルタによって電源周波数を除去した後、バンドパスフィルタによって周波数帯域を 30 - 450 (Hz) に限定する。特徴量の抽出では、表情識別用と可聴化用の 2 種類の RMS を算出する。

窓幅 150 (ms) で算出した RMS を表情識別用として、表情識別部 (3.2 節にて詳述) における学習と識別に用いる。可聴化用 RMS は筋電位信号の細かな変化を捉えるために窓幅 50 (ms) で算出する。フレームシフトについては、表情学習時のみ 1 (ms) とし、表情識別時と可聴化時には 25 (ms) とする。

笑顔表出時には前頭部の筋電位信号にはほとんど変化はなく、側頭部の筋電位信号が変化する傾向がある。そこで、可聴化用 RMS は左右側頭部の筋電位信号から算出する。可聴化用の RMS は算出後正規化を行い、可聴化部で用いる。正規化のための式を式 1 に示す。計測部位 m における中立顔学習時 (3.2 節) の可聴化用 RMS の平均値を R_{0_m} 、笑顔学習時 (3.2 節) の可聴化用 RMS の最大値を R_{max_m} 、 n サンプル目の可聴化用 RMS を $r_m(n)$ とし、 n サンプル目における正規化後の可聴化用 $RMSR_m(n)$ を算出した。

$$R_m(n) = \frac{r_m(n) - R_{0_m}}{R_{max_m} - R_{0_m}} \quad (1)$$

$$R_m(n) = \begin{cases} 1 & (R_m(n) > 1) \\ R_m(n) & (otherwise) \end{cases}$$

3.2 表情識別部

表情識別部では、2 クラス識別器として高い性能を持つ SVM (サポートベクタマシン) を用いて表出されている表情が笑顔かそれ以外の表情かを識別する。本システムでは SVM の表情識別および学習のアルゴリズムとして LIBSVM を用いた [13]。SVM は未学習のデータに対しても汎化能力が高く、計算負荷も大きくないため、リアルタイムでの笑顔識別を行うことが可能である。

SVM で識別を行うためには、事前に学習を行う必要がある。また、表面筋電位信号の強度やパターンは個人差や電極の装着位置によって異なるため、個人ごとに学習を行う必要がある。学習のために、中立顔、微笑み (歯を見せない笑顔)、笑顔 (歯を見せる笑顔)、咬合 (奥歯を噛みしめる) の 4 つの表情をそれぞれ 2 秒間保持してもらう。その間に計測した筋電位信号を用いて個人に応じた信号強度やパターンを学習し、それを元に識別を行う。

3.3 可聴化部

可聴化部では、左右側頭部 2ch の可聴化用 RMS および表情識別結果を入力として、可聴化音を出力する。可聴化音の出力は、表情識別部で笑顔と識別されている間のみ行われる。可聴化の手法の 1 つである Parameter mapping sonification (PMSon) [14] により可聴化音の生成を行う。PMSon はデータを音響パラメータ (音量、音高等) に結びつけること (マッピング) で、データの演歌を音の様子

の変化で表現する手法である。

本可聴化システムの可聴化音への要求として、表情の動きがどのように音に反映されるかが理解しやすい「わかりやすさ」、自発的な笑顔の表出を促す「楽しさ」、長時間聴いていても不快になることのない「心地よさ」の 3 つを設定した。

「わかりやすさ」と「楽しさ」の要求を満たす方法として、音階を用いた音高へのマッピングが有効であると考えた。可聴化用 $RMSR_m(n)$ の値の範囲 (式 1, 0 -1) を用いる音階の音数で等分割し、音階による音高へのマッピングを行った。音階には C メジャーペンタトニックスケールを用いた。表情の動きで音高が変化するため「わかりやすさ」の要求は満たすことができ、また表情の動きに応じて音が細かく変化するため「楽しさ」の要求も満たすことが出来ると考えた。

「心地よさ」の要求を達成するにあたり重要な要素である音色について、人工的な合成音では不適と考え、自然音である泡の音を模した合成音モデルを用いることにした。泡の音の合成アルゴリズムについては書籍「Designing Sound[15]」に記載されている。実際に SuperCollider 上で泡の音の合成アルゴリズムを実装するにあたってはオンライン書籍「Designing Sound in SuperCollider[16]」のうち「Bubbles」(Dan stowell による実装 [17]) を参考にした。泡の音 1 つの持続時間は約 0.1 秒であり、速い立ち上がりその後減衰していく。

実際に可聴化する際は、リズムの面でも変化をつけるために同じ音高が連続した場合は可聴化音は再生されないようになっている。また、自然な泡の音を意識して音高が上がるにつれ音量が小さくなるようにした。

4. 調査 1: 静止画を用いた弱視者の笑顔認識能力調査

弱視者の笑顔認識に対する笑顔可聴化システムの有効性検証にさきがけ、弱視者の他者笑顔認識能力を調査する調査を 2 つ実施した。1 つ目の調査では、静止画を用いて弱視者の他者笑顔認識について調査した。調査は 1 名ずつ行い、全体で 3 名の晴眼者 (S-01 ~ S-03) および 2 名の弱視者 (L-01, L-02) が参加した。弱視参加者の視力、症状及び発症時期を表 1 に示す。

表 1: 弱視参加者の視力、症状及び発症時期 (調査 1)

弱視参加者 id	視力 (左 右)	症状	時期
L-01	0.09 0.4	視野欠損	先天
L-02	0.09 0.09	中心暗点	先天

4.1 実験手順

A4 用紙に印刷された表情画像を 1 枚ずつ椅子に座った

状態の参加者に提示し、その表情が「笑顔かどうか」「回答に自信があるか」をそれぞれ2択で回答してもらった。提示画像は日本人女性の表情データベースである JAFFE データベース [18] のうち、笑顔 10 枚、その他の表情 30 枚の計 40 枚選定した。20 枚を 1 タームとして、休憩を挟んで 2 ターム実施した。

4.2 結果

参加者個人ごとの成績を図 3、晴眼者および弱視者それぞれの正答率および自信度の平均を図 4 に示す。正答率と自信度の算出方法をそれぞれ式 2、式 3 に示す。

$$\text{正答率 (\%)} = \frac{\text{正解数}}{\text{全回答数}} \quad (2)$$

$$\text{自信度 (\%)} = \frac{\text{回答「自信あり」の数}}{\text{全回答数}} \quad (3)$$

また、各ターム及び全体の所要時間の平均を表 2 に示す。弱視者は晴眼者に比べ、回答に自信を持っておらず、判断に時間がかかっている傾向が示された。

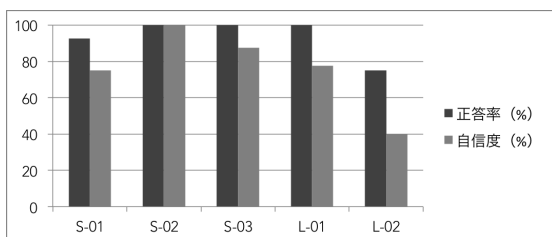


図 3: 個人ごとの成績 (正答率 = 正解数/全回答数, 自信度 = 回答「自信あり」の数/全回答数)

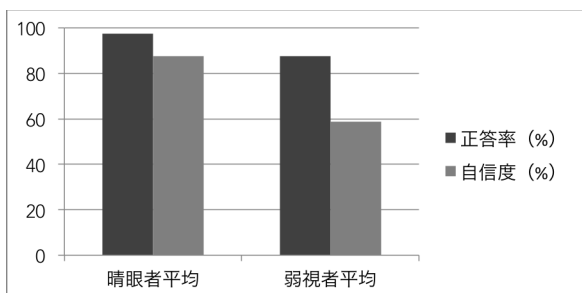


図 4: 晴眼者・弱視者の平均成績 (正答率 = 正解数/全回答数, 自信度 = 回答「自信あり」の数/全回答数)

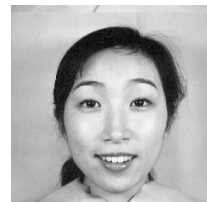
表 2: 平均所要時間

	晴眼者	弱視者
1st	1分11秒	2分11秒
2nd	1分7秒	1分50秒
全体	2分18秒	4分1秒

4.3 考察

弱視者の回答例を図 5 に示す。この例のうち上段左側と右側の表情は笑顔だが、左側の歯が見えている笑顔に対しては両者自信を持って回答しているのに対し、右側の歯が見えていない微笑みに対しては、両者とも回答に自信を持っていない。また、下段の表情は怒りを表しているが、この表情に対しても両者回答に自信を持っていないことに加え、この表情を笑顔と判定した弱視者もいた。考えられる要因として、口角の右側が上がっているため笑顔だと判断された可能性が考えられる。終了後に弱視参加者に話を伺うと、「歯が見えているとわかりやすいが、見えていないと難しい」といったコメントがあり、微笑みなどの度合いの小さい笑顔の認識は難しい可能性が示唆された。

また、弱視者は晴眼者に比べ、回答に自信を持っておらず、判断に時間がかかっている傾向があるため、リアルタイムで表出される笑顔に対してはさらに認識が困難であると考え、動画を用いた調査を実施することにした。



(a) 笑顔
両者正解
両者自信あり



(b) 笑顔
両者正解
両者自信なし



(c) 笑顔ではない
1名正解 1名不正解
両者自信なし

図 5: 弱視者回答例

5. 調査 2: 動画を用いた弱視者の笑顔認識能力調査

静止画を用いた調査の結果を踏まえ、映像を用いて弱視者による他者の笑顔認識について調査した。調査は 1 名ずつ行い、全体で 3 名の晴眼者 (S-01 ~ S-03) および 4 名の弱視者 (L-01 ~ L-04) が参加した。弱視参加者の視力、症状及び発症時期を表 3 に示す。

5.1 実験手順

参加者は椅子に座った状態で PC の画面上に表示される

表 3: 弱視参加者の視力、症状及び発症時期 (調査 2)

弱視参加者 id	視力 (左 右)	症状	時期
L-01	0.2 0.08	視野欠損	先天
L-02	0.01 0.02	中心暗点・色覚異常	高校 1 年
L-03	0.1 0.08	視野欠損・暗順応	先天
L-04	0.4 0.7	眼振	先天

映像を 2 回視聴する。映像は映画作品中から抽出した。刺激に用いた映画作品のタイトルと切り出した時間を表 4 に示す。今回の調査では視覚情報のみでの弱視者の笑顔認識を調査するため、刺激に音声は含めなかった。

参加者は 2 回の視聴のうち、1 回目の視聴で映像中どのタイミングに笑顔が出現するか、あるいは映像を通して出現しないかを把握し、2 回目の視聴で実際にキーボードで回答した。回答として、映像中に笑顔が出現したと思うタイミングで 'f' キーを、笑顔が消失したと思うタイミングで 'j' キーを押下する。笑顔が出現していないと思う場合はキーを押下しない。この 2 回の視聴を笑顔を含む映像に 6 種類に加え笑顔を含まない映像 3 種類の合計 9 種類の映像に対して実施した。

表 4: 用いた映画作品のタイトルおよび切り出し時間

映像 id	映画タイトル	切り出し時間	笑顔の有無
1	朝食亭	47:00~47:18	無し
2	朝食亭	78:30~78:55	無し
3	朝食亭	85:40~85:58	無し
4	船を編む	49:40~50:11	有り
5	船を編む	71:10~71:28	有り
6	グーグーだって猫である	49:39~50:00	有り
7	しあわせのパン	1:20~1:50	有り
8	しあわせのパン	32:00~32:27	有り
9	ソラニン	53:39~54:09	有り

5.2 結果

ここでは特に顕著な結果が見られた id6 と id7 (表 4) の 2 つの映像について述べる。id6 の映像に対しての各参加者の回答を図 6 に、id7 の映像に対しての各参加者の回答を図 7 に示す。横軸は時間を表しており、薄いグレーのエリアが晴眼者が笑顔だと判定した部分、濃いグレーのエリアが弱視者が笑顔と判定した部分を表している。また、id6, 7 以外の笑顔有り映像については晴眼者間でも回答にばらつきがあり、顕著な結果は得られなかった。

5.3 考察

id6 の映像に対して、晴眼者は全員 21 秒付近で笑顔が消失したと判断しているが、弱視者のうち 2 名は 18 秒付近で笑顔が消失したと判断している。実際の映像では 16 秒付近から女性が微笑み、18 秒付近で微笑みのまま顔の角度

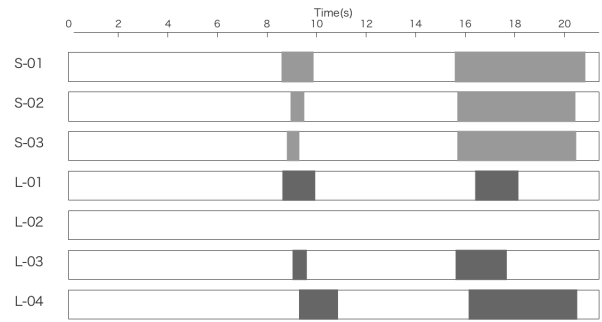


図 6: id6 に対する各参加者の回答結果 (色付き部分: 参加者が笑顔だと回答した区間, 薄いグレー: 晴眼者 濃いグレー: 弱視者)

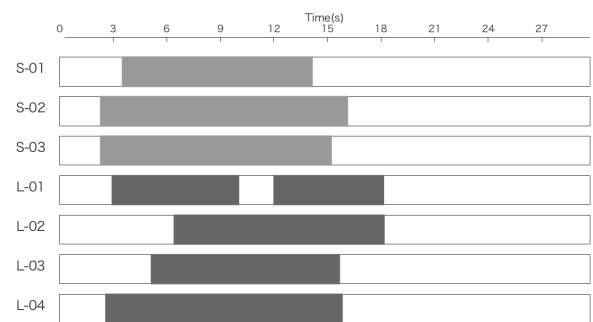


図 7: id7 に対する各参加者の回答結果 (色付き部分: 参加者が笑顔だと回答した区間, 薄いグレー: 晴眼者 濃いグレー: 弱視者)

を正面から斜めに変えている。このことから、弱視者は正面以外の角度で表出される笑顔の認識が難しい可能性が示唆された。

また、id7 の映像に対して、晴眼者は全員 3 秒付近で笑顔が出現したと判断しているが、弱視者のうち 2 名は 2 秒ほど遅れて笑顔が出現したと判断している。実際の映像では 3 秒付近で女性が 2 秒ほど微笑み、その後笑顔を表出するシーンになっている。笑顔出現の判定が遅れた弱視者は、最初の微笑みを認識することが出来ておらず、笑顔になり歯が見えたタイミングで笑顔が出現したと判断している。このことから、弱視者は笑顔を判断する際に歯が見えているかどうかを大きな判断基準にしている可能性が示唆された。

6. おわりに

本研究では、笑顔可聴化システムを用いて弱視者による他者の笑顔認識の支援を目指している。笑顔可聴化システムは顔面の表面筋電位信号を入力として、笑顔識別を行い笑顔時に可聴化音をリアルタイムに出力する。

笑顔可聴化システムの有効性検証にさきがけ、弱視者の

他者の笑顔認識能力について静止画と動画を用いる2種類の調査によって調べた。静止画を用いた調査より、弱視者は歯が見えない微笑みのような度合いの小さい笑顔の認識が難しく、またリアルタイムに表出された表情の判断に困難がある可能性が示唆された。動画を用いた調査より、弱視者は笑顔の判断に歯が見えているかどうかを大きな基準にしている可能性、正面以外の角度で表出される笑顔の認識が難しい可能性が示唆された。

以上を踏まえ、笑顔可聴化システムを用いることで今回の調査より認識が難しい可能性が示唆された歯の見えない微笑みや正面以外の角度で表出される笑顔に対する認識の支援が可能であると仮説を立て、今後検証実験を行う。

参考文献

- [1] 厚生労働省社会・援護局障害保健福祉部企画課：平成18年身体障害児・者実態調査結果(2008).
- [2] 大山正, 今井省吾, 和氣典二, 菊池正：新編 感覚・知覚心理学ハンドブック Part2 (2007).
- [3] 中野泰志, 相羽大輔, 小松真也：ロービジョンの表情認知を促す方策—対面コミュニケーションの課題と工夫に関する実態調査からの考察—, 日本視能訓練士協会誌, pp. 55–63 (2014).
- [4] Kramer, G., Walker, B. and Bargar, R.: Sonification Report: Status of the Field and Research Agenda (1999).
- [5] Ekman, P.: An argument for basic emotions, *Cogn. & Emot.*, Vol. 6, No. 3-4, pp. 169–200 (1992).
- [6] Russell, J. A.: A circumplex model of affect., *J. Per. Soc Psychol.*, Vol. 39, No. 6, p. 1161 (1980).
- [7] 木塚朝博：表面筋電図, Practical usage of surface electromyogram, 東京電機大学出版局, 東京, Japan (2006).
- [8] Patil, V., Akhtar, M. Q., Parab, A. and Fernandes, A.: Sonification of Facial Expression Using Dense Optical Flow on Segmented Facial Plane, *International Conference on Computing and Control Engineering (ICCCCE)* (2012).
- [9] Funk, M., Kuwabara, K. and Lyons, M. J.: Sonification of facial actions for musical expression, *Proceedings of the 2005 Conf. NIME.*, National University of Singapore, pp. 127–131 (2005).
- [10] Takano, Y. and Suzuki, K.: Affective Communication Aid Using Wearable Devices Based on Biosignals, *Proc. of the 2014 Conf. Interact. Design and Children (IDC2014)*, New York, NY, USA, ACM, pp. 213–216 (online), DOI: 10.1145/2593968.2610455 (2014).
- [11] : "http://opensoundcontrol.org".
- [12] Phinyomark, A., Nuidod, A., Phukpattaranont, P. and Limsakul, C.: Feature extraction and reduction of wavelet transform coefficients for EMG pattern classification, *Elektronika ir Elektrotechnika*, Vol. 122, No. 6, pp. 27–32 (2012).
- [13] : "http://www.csie.ntu.edu.tw/~cjlin/libsvm".
- [14] Hermann, T., Hunt, A. and Neuhoff, J. G.(eds.): *The Sonification Handbook*, Logos Publishing House, Berlin, G (2011).
- [15] Farnell, A.: *Designing Sound*, The MIT Press, Cambridge, Massachusetts (2010).
- [16] : "http://en.wikibooks.org/wiki/Designing_Sound_in_SuperCollider".
- [17] : "http://www.mcl.d.co.uk/research/#phd".
- [18] Lyons, M. J., Akemastu, S., Kamachi, M. and Gyoba,