

プレゼン先生：音声情報処理と画像情報処理を用いたプレゼンテーションのトレーニングシステム

A Presentation Training System using Speech and Image Processing

栗原 一貴 後藤 真孝 緒方 淳 松坂 要佐 五十嵐 健夫*

概要. 本論文では、音声情報処理と画像情報処理を組み合わせたプレゼンテーショントレーニングシステムを提案する。開発したシステム「プレゼン先生」は、プレゼンテーションツールと連携しつつ、マイクおよび Web カメラから得られた発表者の音声および振る舞いを分析し、話速度、声の抑揚、聴衆とのアイコンタクトの度合いなどの指標をリアルタイムに発表者にフィードバックする。また、これらの指標がある閾値を超えた際に警告を通知する。発表終了後には、これらの指標をグラフとして可視化して提示することで、発表者の反省作業を支援する。本システムは、プレゼンテーションの内容そのものの改善ではなく、発表中に意図せず行ってしまう不適切な行動を抑制することにより、発表者単独によるプレゼンテーションスキルの底上げを行うことができる。小規模なユーザテストにより、本システムの有効性が示され、更なる改善に向けてのコメントが得られた。

1 はじめに

プレゼンテーションは現代において重要な役割を果たしている。書店には方法論を教える書籍があふれており、社会のプレゼンテーションに対する関心の高さが伺える。これらの教本では、主に発表に至るまでの準備段階における発表戦略の練りおおよび資料のまとめ方と、発表中における話術および身体的所作のあり方を論じている [19]。この両者のうち、発表戦略および資料のまとめ方については発表者が準備段階に繰り返し推敲して質を高めることが比較的容易なのに対し、話術および身体的所作については実際に発表全体を通してのリハーサルを行うことでしか現状分析を行うことができない。発表者自身が「評価者の視点」を持ちつつ発表を行いにくいいため、ビデオ撮影環境もしくは助言者の存在なくしては改善策の検討を行うことが従来難しかった。

そこで我々は、音声情報処理と画像情報処理を組み合わせてプレゼンテーションリハーサルを自動的に評価し、発表者にリアルタイムもしくは事後フィードバックを行うことで自己トレーニングが可能なシステム「プレゼン先生」を開発している。無論、プレゼンテーションの総合的な評価を完全に機械が判断することは難しい。プレゼン先生システムは、その総合的な評価を構成する多様な指標のうち、発表者が意図せず行っていてマイナス評価になりうるいくつかの項目を排除することで、プレゼンテーションスキルの底上げを図るものである。

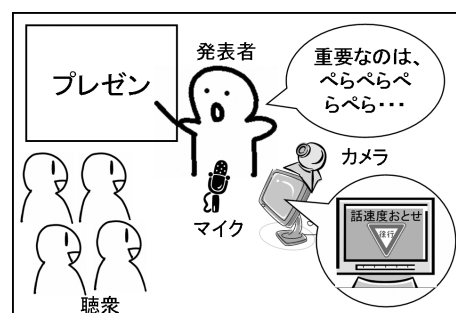


図 1. プレゼン先生システム

一般にプレゼンテーションでは、適切な速度で快活に話すこと、言い淀みが少ないこと、視線が主に聴衆の方に向いていること、適切な「間」を設けること、時間配分が適切であること、などが好ましい話術および身体的所作として挙げられている。これらは、マイクで音声を録音し音声情報処理を行うことにより、またカメラで話者を撮影し画像情報処理を行うことにより検出が可能である。プレゼン先生は、プレゼンテーションツールと連携しながら発表練習中にこれらの指標をリアルタイムに数値化して表示するとともに、好ましくない状態に陥った場合に発表者にアラートを通知する（図 1）。また、発表終了後に発表全体の評価をグラフ化し、発表者に反省の機会を提供する。

音声情報処理や画像情報処理においては、予期せぬ入力に対するロバスト性を達成することが難しい点がしばしば問題になる。しかしプレゼンテーションリハーサルという行為は実行場所や使用機材などの環境を比較的自由に設定できる特徴をもつ。よって静かな場所で、カメラ画像には発表者のみが映り背景が変化しないなど、既存技術に適した環境においてシステムを利用することを前提とすることが現

Copyright is held by the author(s).

* Kazutaka Kurihara and Takeo Igarashi, 東京大学大学院 情報理工学系研究科 コンピュータ科学専攻, Masataka Goto and Jun Ogata and Yosuke Matsusaka, 産業技術総合研究所, Takeo Igarashi, 科学技術振興機構 さきがけ

実的であり、高い実用性を実現することができる。

2 関連研究

本論文はリハーサル支援という、プレゼンテーションツールの一つの拡張機能についての報告である。プレゼンテーションの編集・発表機能に関する報告は数多いが、スライドごとの発表時間をフィードバックする PowerPoint [3] のリハーサル機能を除いて、リハーサル支援を扱ったものは調査した限りでは存在しない。

本研究はプレゼンテーションにおける、発表内容以外の情報、即ち非言語情報を取り扱う。非言語情報を取り扱った関連研究として、後藤ら [16] による「音声補完シリーズ」が挙げられる。これはこれまで音声認識技術において活用されていなかった非言語情報を積極的に取り入れた音声インタフェースである。一方 Mehrabian [10] は、対人コミュニケーションにおいて感情を伝える場合に、言語情報が寄与するのはわずか 7% に過ぎず、残りの 93% は声や表情などの非言語情報によって決定されたという実験結果を報告した。この報告 [10] に基づき非言語情報を意識的に活用し、また意図しない非言語情報伝達を抑制することの重要性を説く一般書籍が出版されている [19][18]。本研究はそれを支援するシステムを構築する。

本研究では、音声認識、画像認識などを組み合わせたマルチモーダルシステムを取り扱う。特に、コンピュータへの入力のためにユーザが行った発話や身体動作を認識・活用するのではなく、ある別の目的のためにユーザが自然に行った発話や身体動作を観察し二次利用する。このテーマについては Hindus ら [7]、Lyons ら [9]、栗原ら [8] が取り組んでいるが、我々はこれをプレゼンテーションリハーサルへと応用する。

認識技術をトレーニングに活用した例として、外国語学習支援を行う [14] および TalkMan [4]、役者の気分を楽しむしばいみち [5] などが挙げられる。また、[6] は「コンピュータが人間に指示する・指導する」場合にどのようなインタラクションデザインをすべきかを論じている。

3 プレゼンテーションの評価指標

文献 [19][18] によると、プレゼンテーションの良し悪しに影響を与える指標として、話速度、発話の抑揚、言い淀み、聴衆とのアイコンタクトの度合い、「間」の使い方、発表時間の遵守などが挙げられている。我々は各種文献を参考に、これらに対する標準値と異常値の閾値を決定した。

プレゼンテーション中に普通の人は早口になる傾向にあるが、実際は普段より少しゆっくり話した方がよい [19]。一方、抑揚のないモノトーンな発話も聴き手を退屈させてしまう要因となる [19]。そこで

一定の話速度（単位時間あたりの発話音節数¹）を超えたとき、また抑揚の指標として声の高さ（F0）の標準偏差が一定以下になったときに警告を通知することを考える。話速度と、声の高さおよびその標準偏差により、様々な音声資料を分析したものとして郡による報告が挙げられる [11]。これを参考に我々は話速度の上限値を 7.6 音節/秒、F0 標準偏差の下限値を 10Hz（男性の場合）と定めた。

言い淀みの存在も、意図せずプレゼンテーションのパフォーマンスに影響を与えてしまう要素の一つである [19]。発表者が「えー」のような有声休止（母音の引き伸ばし）による言い淀みをおこなった場合は、直ちに警告を通知する。

以上は音響的な評価指標であったが、視覚的な評価指標として、聴衆とのアイコンタクトの度合いが上げられる。文献 [19] によると発表者が聴衆の方を向いている時間的割合をアイコンタクト率と定義する時、アイコンタクト率が 15% 以下になると、話し手は冷たい、弁解的、未熟などの印象を与えるという。一方 80% 程度であれば、自信がある、誠実、親近感、熟練などの印象を与えるという。我々は 15% 基準を下回った場合に発表者に警告を通知するデザインとした。

音響的な特徴と視覚的な特徴を同時に考慮すべきマルチモーダルな指標として、「間」の取り方が挙げられる。プレゼンテーションにおける間とは、故意に沈黙を作り出すことある。これにはその後の発言を強調したり、聴衆を話に引き込む効果がある [19][18]。この場合、沈黙とは単に発話していないだけでなく、聴衆の方を向くことで意識的にその無音区間を伝えようとしている姿勢を示すことが必要である。文献 [18] によると、1～2 秒程度を通常の間、5 秒以上を「びっくり間」として使い分けている例が挙げられている。逆に間がなく、一つ一つの発話区間が冗長であることは聴衆の理解を妨げる。これらを考慮し、我々は発表者が間の取り方を練習するために、現在の間の様子を表す指標 SI (Speech Index) を以下のように定義する。

無音区間：連続して聴衆を見ている無音区間を t (秒) として

$$SI = 50 \quad (\text{when } t < 1)$$

$$SI = 50 + 12.5(t - 1) \quad (\text{when } t \geq 1)$$

$$SI = 100 \quad (\text{when } SI > 100)$$

発話区間：連続した発話時間を t (秒) として

$$SI = 50 - 50t/13$$

$$SI = 0 \quad (\text{when } SI < 0)$$

SI は 50 を基準値として、間を取っていると大きくなり、「びっくり間」である 5 秒で最大値 100 とな

¹ 音節とは日本語における「かな」に対応する音韻体系を指す。

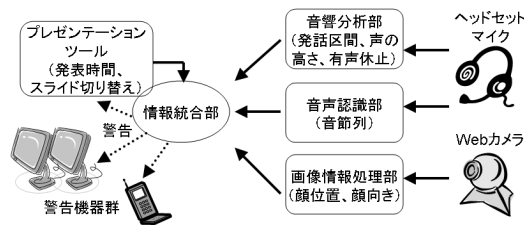


図 2. プレゼン先生のシステム構成



図 3. 視覚による情報提示の例

る。一方発話を続けていると 50 から徐々に低下し、13 秒喋り続けると最小値 0 となる。この 13 秒と言う値は、深呼吸の時間を参考に決定した。間の取り方については重要性は多く指摘されているものの、具体的な良し悪しの基準は調査した限りでは存在しないため、現時点では閾値を設けて警告を通知することは行わない。

最後に、発表時間を遵守することも大切である。全発表時間の 80% と 100% が経過した時、発表者に通知することとする。

4 プレゼン先生システム

前章で定義した評価指標をもとに、プレゼンテーションリハーサルを監督し発表者に改善を促すトレーニングシステム「プレゼン先生」のプロトタイプを開発した。以下にその詳細を示す。

4.1 システム構成

プレゼン先生システムは、プレゼンテーションツール「ことだま」[13]の拡張機能として実装されている。システムの構成図を図 2 に示す。音響分析部および音声認識部はマイク音声から発話区間認識、F0 推定、有声休止検出、音節単位の音声認識を行い、結果を常に情報統合部に送信している。画像情報処理部はカメラ画像から得られた発表者の顔の位置および顔の向きの情報を常に情報統合部に送信している。情報統合部は音声、画像情報に加えてプレゼンテーションツールから監督開始および終了の信号、発表時間信号、スライド切り替えの信号およびスライドのサムネイル画像を受け取り、不特定多数の警告機器群を介してユーザにフィードバックする。これらの間の通信は RVCP プロトコル [16] で行っており、すべてのシステム構成要素をローカルネットワーク上の任意のホストで実行させることが可能であるため、負荷分散が容易である。

4.2 ユーザインタフェース

プレゼン先生システムはユーザの発表をもとに、情報統合部で最終的に算出された各評価指標のフィードバックを行う。フィードバックには、リアルタイムの「オンライン」と事後の「オフライン」の 2 種類を用意した。

4.2.1 オンラインフィードバック

オンラインフィードバックは発表中に逐次的に短期的な統計値を通知するものである。リアルタイムモニタ(図 3 左)は、統計量をそのまま表示し続けるもので、最近のパフォーマンスを視覚的に確認するために用いられる。またアラートは閾値を超えた統計量について、ユーザに図 3 右のような 6 種類の通知を行うものである。アラートは主に画面に視覚的に提示され、画面の右下からポップアップする小型のもの、全画面表示されるものの 2 種類がある。これら視覚による情報提示は、発表練習環境に自由に設置された各種ディスプレイ(メインスクリーン、原稿表示用画面、警告表示用サブ画面²⁾)ごとに提示項目を選択できる。アラートには視覚のほかにも音および音声で通知するもの、振動で通知するもの³⁾を実装しており、ユーザがアラートの種類に応じてモダリティを選択できる。

4.2.2 オフラインフィードバック

オフラインフィードバックは発表後に発表全体を振り返り、蓄えられたデータをグラフにして発表者に提示するものである(図 4)。様々な指標に関する時系列データは時刻情報およびその時用いていたスライドと対応付けられ、スライドサムネイル画像とともに色分けされて表示される。それぞれの指標について、逐次的な統計量をプロットしたものとスライド単位で統計量の平均値を求めプロットしたものの 2 種類のグラフが作成される。このため、全体的に見てどの辺りを改善すべきか、というマクロな判断と、特定の箇所におけるミクロな判断の両方が可能である。

システムが作成した多数のグラフ群は、プレゼンテーション資料ファイルに蓄積される。利用したプレゼンテーションツールであることだまは、編集モードと発表モードの区別がなく、一枚の広い 2 次元模造紙上に発表資料を作成し、事前登録された特定の矩形領域をスライドのように移動することによりプレゼンテーションを行う。よって提示する資料の周

²⁾ 基本的に原稿を見ずに聴衆の方向を見ることが望ましいため、聴衆の位置にサブ画面を設置し警告表示を行うことが効果的である。

³⁾ 現在のプロトタイプ版では、マナーモードにした携帯電話に対しメールを送信することで実現している

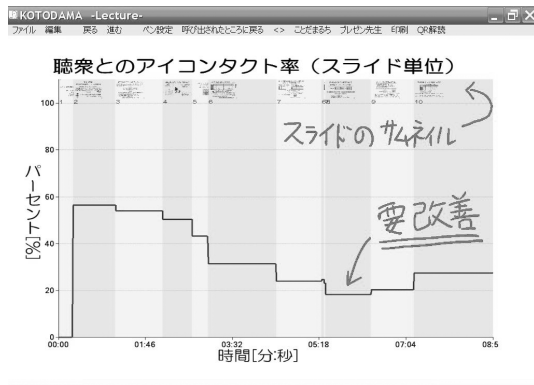


図 4. オフラインフィードバックとして作成される個々のグラフの例：ペンによるアノテーションが可能。

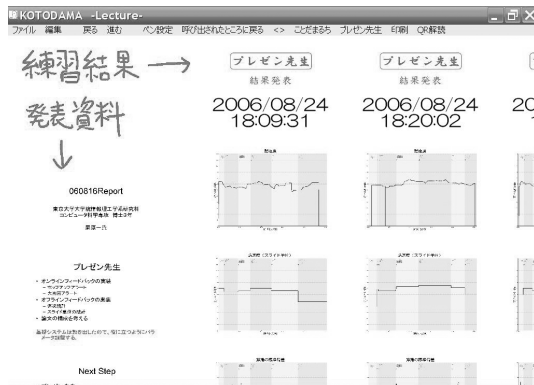


図 5. プレゼンテーション資料に蓄積される過去の発表練習履歴

囲の領域に、発表に関連する情報を自由に書き加えることが可能である。一度発表練習を行うと、資料の右端にグラフ群がタイムスタンプとともに添付され、閲覧やペンアノテーションによるコメント書き込みが行える。また左右方向には過去の練習における同一指標の履歴が残っているため、比較も容易である(図5)。

オフラインフィードバックについて、現在はグラフの提示のみにとどまっているが、今後多数の発表データを収集し、それと比較しての総合評価、改善に向けてのアドバイス提示などを行う予定である。

4.3 音響分析部(発話区間, 声の高さ, 有声休止)

音響分析部では、マイクから得られた発表者の音声を入力として発話区間, 声の高さ(F0), 有声休止を10msごとにリアルタイムに求め、情報統合部に送信する。発話区間は、音声のパワーの大きい箇所に基づいて検出する。F0推定と有声休止検出には、文献[15]のF0推定手法、有声休止検出手法を用いる。F0推定手法は、背景雑音等を伴う音響信号に対しても口バーストに機能する特長を持ち、コムフィルタの考え方に基づいて、最も優勢な高調波構造のF0を音声のF0として推定する。一方、有声休止の

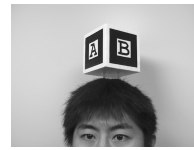


図 6. AR ツールキットを用いたマーカ

検出手法は、任意の母音の引き延ばしを言語非依存に検出できる特長を持ち、有声休止が持つ二つの音響的特徴(F0の変動が小さい、スペクトル包絡の変形が小さい)をボトムアップな信号処理によって検出する。

4.4 音声認識部(話速度)

音声認識部では、マイクから得られた発表者の発話を入力として、音節を単位とした音声認識を行い、その認識結果(音節列)と対応する発話区間情報を情報統合部に送信する。検出された音節数を発話区間の長さで割ることで、無音区間を除く単位時間当たりの音節数を計算する。

音声認識器としては、julian [2]を、各発話の認識結果が情報統合部に逐次送信されるように拡張したものをを用いた[17]。認識時の言語モデルとしては、121種類の音節(無音も含む)が任意の接続を許すネットワーク文法を用いている。

本手法による音声認識は通常スペックのラップトップPCを用いても数秒程度の遅れで出力される。これはユーザが発表中に最近の話速度をチェックするという用途には十分な性能である。

4.5 画像情報処理部(顔の位置と向き)

画像情報処理部では、Webカメラから取得した画像から発表者の顔の位置および顔の向きを計算し、情報統合部に逐次送信する作業を行う。プロトタイプ版の実装として、ARツールキット[1]を用いたマーカによる方式と、部分空間法およびSVM(Support Vector Machine)を用いた純粋な画像処理による方式[12]の2方式を開発した。どちらの手法も単眼カメラからリアルタイムに発表者の顔の位置と向きの6自由度情報を得ることが可能である。以下に詳細を示す。

4.5.1 AR ツールキットを用いたマーカ方式

ARツールキットを用いたマーカ方式では発表者は図6で示すような特殊なマーカを頭部に着用する。発泡スチロール製の立方体の各面にARツールキットの2次元コードを貼ることにより、発表者がどこを向いても頭部の位置と向きの検出を行うことが可能である。本方式はユーザごとに画像処理用の学習データ等を与えなくても動作する。



図 7. 部分空間法と SVM を用いた 2 次元画像からの 360 度方向顔・顔部品追跡手法

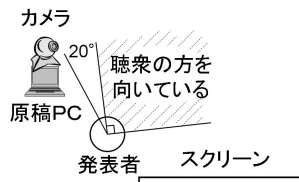


図 8. カメラ、スクリーン、発表者と聴衆の方向の定義：スクリーンに向かって右側に演台があり，Web カメラが原稿 PC の上に設置されていることを想定．

4.5.2 部分空間法と SVM を用いた画像処理方式

部分空間法と SVM を用いた画像処理方式では，まず事前準備として発表者のあらゆる姿勢における顔領域画像をデータとして収集し，それらデータに対して主成分分析を適用することで固有ベクトルのセットを得る．それら固有ベクトルのセットをモデルとして用い，入力画像に対して最もフィットするモデルを判別することで様々な姿勢に適応性のある顔位置の追跡を行うことができる．また検出された顔位置に対して SVM を用いた顔角度推定を適用することで顔の向きも同時に求める（図 7）．本手法では発表者はマーカを頭部に着用する必要がなく，自然な発表が可能になる．

4.5.3 聴衆を見ている状態の定義

先述の 2 方式のいずれかを用いて発表者の顔の位置と向きが得られた後，それをもとに発表者が聴衆を見ているか否かの 2 値情報に変換する．プロトタイプ版では実装の簡単のために，図 8 のようなリハーサル環境を仮定し，顔の向きの水平方向角度が一定範囲内のときに聴衆を見ているとみなすアルゴリズムを採用している．今後はシステム利用前に環境のキャリブレーションを行うことにより，より柔軟な状態検出を行う予定である．

5 ユーザスタディ

実装したプレゼン先生のプロトタイプを用いて，性能評価とユーザフィードバックを得るための小規模なユーザスタディを行った．

5.1 方法

プレゼンテーションの経験がある 3 人の男性大学院生 A, B, C を被験者として，彼らが過去に行ったことのあるプレゼンテーションを本システム上で

8 分間行わせた⁴．画像処理法としては AR ツールキットのマーカ方式を採用し，視覚によるオンラインフィードバックのみを行った．終了後，オフラインフィードバックのグラフ群を用いて発表を振り返ってもらい，本システムの有効性，改善点についてアンケート記入を求めた．

また，システムの出力が適切であったかどうかを検証した．被験者の発表の様子をシステム外部からビデオ撮影し，著者・被験者とは異なる第三者 1 名が手作業で分析した．言い淀み数およびアイコンタクト率については直接値を求め，話速度および抑揚については分析者の主観的な印象を記録した．

5.2 結果

5.2.1 システムの有効性，改善点

3 人の被験者は自己の発表中の状態が数値化され，今まで気づけなかった自己の挙動を把握することができた点を評価した．またシステムの各評価指標，警告について，概ね役に立つとコメントした．一点，発表残り時間については全体の残り時間ではなく，予め設定した各スライドごとの予定時間に対してアラートの方が有効であるという意見があった．

5.2.2 システムの判断と人間の判断の比較

被験者が実際に行った言い淀みは被験者 A, B, C でそれぞれ 18 回，73 回，1 回で，それに対しシステムが検出に成功し警告を行った回数は 1 回，12 回，0 回であった．一方実際に言いよどみの無い場所で誤って警告を行ったことはなかった．システムが検出できなかった言い淀みは主にごく短い時間しか継続しない種類のものである．現在は再現率よりも適合率を重視した設定となっているが，ユーザの希望によりパラメータを調整することを今後考えたい．

図 9 に，被験者のアイコンタクト率についてシステムの出力と分析者の手作業評価結果を示す．被験者 A および C については両者は概ね一致した．被験者 B については，スクリーンと聴衆の丁度中間の方向を向いて話す傾向があったため，検出アルゴリズムが有効に働かず一致度は比較的低かった．

表 1 に話速度と抑揚に対する評価者の印象と，システムが行った警告の対応を示す．話速度に関しては人間の判断とシステムの判断は相関が高いが，抑揚については警告の閾値が低すぎた．3 人の被験者の F0 標準偏差について発表を通して分析した結果，20Hz ~ 40Hz 程度に閾値を設定すれば分析者の主観に即した警告を行えたということがわかった．現在は Hz 単位に基づく基準を設定しているが，今後幅広い声域のユーザへの対応を考え，対数周波数による F0 標準偏差の評価を検討中である．

今回の実験では，全体的に警告の基準は控えめで

⁴ PowerPoint で作成した資料をインポートする機能を用いた．

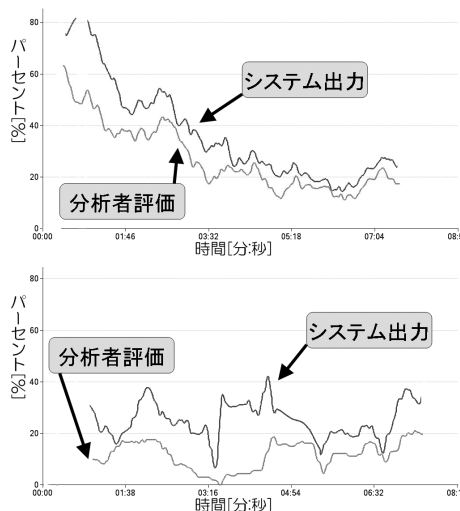


図 9. アイコンタクト率の比較：(上) 一致度の高い被験者 C (下) 一致度の低い被験者 B。

ある。しかし被験者らは、システムに「監督されている」という状態にあること自体が自己の不適切行為にブレーキをかけるプラスの効果をもつとコメントしている。

表 1. 話速度、抑揚に関する分析者の印象とシステムの警告頻度の対照

	話速度 印象	話速度 警告頻度	抑揚 印象	抑揚 警告頻度
被験者 A	速すぎ	頻繁	快活	なし
被験者 B	速すぎ	頻繁	快活	なし
被験者 C	丁度良い	なし	単調	なし

6 まとめと今後の課題

本論文では、音声情報処理と画像情報処理を用いたプレゼンテーションのトレーニングシステム「プレゼン先生」のプロトタイプを開発し、小規模なユーザスタディにより性能評価とユーザからのフィードバックを得た。

今後の課題として挙げられるのは、主に以下の3点である。第1に、リアルタイムフィードバックについて、どのようなモダリティで、どの程度アラートを発表者に伝えることが効果的かを明らかにすることである。第2に、改善を促すアラートを発生させる閾値の設定方法である。現在の実装では著者らが文献を参考に経験的な値を設定しているため、一般性が乏しい点が問題である。視線方向検出のキャリブレーション手法およびユーザの好みに合わせて各閾値を設定できる機能を備えるとともに、巧拙多様なプレゼンテーション映像をもとに最適値を決定する手法の開発が望まれる。第3に、今回は用いなかったプレゼンテーションの評価指標を新たに導入することである。たとえば、視線や立ち位置がふらふらせず安定していることはプレゼンテーションの印象に影響を与えるだろう。また、身振り手振りな

どのボディジェスチャーや表情などのより高度な身体表現が検出できれば、さらに詳細なトレーニングが可能となる。発表資料中のどこを見ているかという情報を用いる分析も興味深い。

謝辞

本研究は、科学研究費補助金（平成 17 年度 若手研究 (B)17700095、平成 18 年度 特別研究員奨励費 18・11190）の援助によって実施された。

参考文献

- [1] AR-toolkit. <http://www.hitl.washington.edu/artoolkit/>.
- [2] julian. <http://julius.sourceforge.jp>.
- [3] PowerPoint. <http://www.microsoft.com/office/powerpoint/prodinfo/>.
- [4] TalkMan. <http://www.jp.playstation.com/scej/title/talkman/>.
- [5] しばいみち. <http://www.jp.playstation.com/scej/title/shibaimichi/index.html>.
- [6] Heer et al. Presiding Over Accidents: System Mediation of Human Action. In *CHI'04*, pp. 463–470, 2004.
- [7] Hindus et al. Ubiquitous Audio: Capturing Spontaneous Collaboration. In *CSCW'02*, pp. 210–217, 1992.
- [8] Kurihara et al. Speech Pen: Predictive Handwriting based on Ambient Multimodal Recognition. In *CHI'06*, pp. 851–860, 2006.
- [9] Lyons et al. Augmenting Conversations Using Dual-Purpose Speech. In *UIST'02*, pp. 237–246, 2004.
- [10] A. Mehrabian. Silent messages, Implicit Communication of Emotions and Attitudes. In *2nd Ed.*, Wadsworth Pub. Co., 1981.
- [11] 郡士郎. アナウンスやナレーションに見られるスタイルの音響的特徴. 第 16 回日本音声学会全国大会予稿集, pp. 151–156, 2002.
- [12] 松坂要佐. 部分空間法と SVM を用いた 2 次元画像からの 360 度顔・顔部品追跡手法. 信学技報 PRMU vol.106, no.72, pp. 19–24, 2006.
- [13] 栗原 他. 編集と発表を電子ペンで統一に行うプレゼンテーションツールとその教育現場への応用. コンピュータソフトウェア, (印刷中).
- [14] 五十里 他. ユーザー発話のセグメンテーションと発話評価機能をもつ英語学習支援システム. 情報処理学会研究報告, SLP40-02, pp. 7–12, 2002.
- [15] 後藤 他. 自然発話中の有声休止箇所のリアルタイム検出システム. 電子情報通信学会論文誌 D-II, Vol.J83-D-II, No.11, pp. 2330–2340, 2000.
- [16] 後藤 他. 音声補完: 音声入力インタフェースへの新しいモダリティの導入. コンピュータソフトウェア, Vol.19, No.4, pp. 10–21, 2002.
- [17] 北山 他. 音声スタータ: “SWITCH” on Speech. 情報処理学会 音声言語情報処理研究会 研究報告 2003-SLP-46-12, Vol.2003, No.58, pp. 67–72, 2003.
- [18] 竹内一郎. 人は見た目が 9 割. 新潮新書, 2005.
- [19] 八幡純芦史. パーフェクトプレゼンテーション. 生産性出版, 1998.