発表練習支援システムのためのプレゼン映像からの癖除去機構の開発

松浦 辰雄　大囿 忠親　新谷 虎松

多くの聴講者の前で発表を行う際に，緊張や不安から声が小さくなったり，口癖などの癖が出てしまうことがある．発表中の癖により聴講者は発表内容が聞き取り辛かったり，集中できないことがある．そのため発表者は，より良い発表を行うために事前に発表練習を行うが，無意識な癖により自身の癖に気づけなかったり，発表中に客観的に自分の発表を評価し癖を矯正するのは困難である．そこで，本研究では，発表している発表者の様子をプレゼン映像として記録し，発表者の癖の検出を行う．検出した癖の箇所や回数を可視化し，発表者は客観的に自分の発表の評価を行うことができる．さらに，過去のプレゼン映像から癖を削除したプレゼン映像を生成し，発表者がプレゼン映像を模倣することで癖の矯正を図る．本稿では過去のプレゼン映像から発表者の癖を検出し，癖の除去を行う癖除去機構について述べる．

A presenter ' s bad habits caused by his/her tension and anxiety may affect the presentation, and audience cannot concentrate their attention to the presentation. Therefore, we practice making presentations beforehand to make better presentations. However, it is hard for us to notice unconscious bad habits and to evaluate ourselves objectively. In this study, we develop a system to detect our bad habits from presentation videos. The system enables users to evaluate their presentations objectively by visualizing occurrence of bad habits. Moreover, the system removes bad habits from presentation movies, and it helps users to correct bad habits by imitating the movies. This paper shows that a habit less presentation movie generator detects and deletes bad habits.

1　はじめに

学会発表等におけるプレゼンテーション（以降，プレゼン）の練習を支援するシステムが研究されているが，発表中の癖の矯正に関する支援が不十分である．本研究では，プレゼン中の癖の矯正を目的とした模範映像の生成を行う．本稿では，プレゼン映像からの癖の検出および除去を実現するための手法について述べる．プレゼン中の発表者の癖が，聴講の妨げとなることがある．発表者は多くの聴講者の前で緊張や不安から声が小さくなったり，口癖などの癖が出てしまう場合がある．それにより，聴講者は，発表者の癖により発表内容が聞き取り辛かったり，集中できないことがある．

発表練習中に自分の癖を自覚できない場合がある．先行研究では，発表者が発表している様子を記録した発表履歴から発表の癖を検出し癖の出ている箇所や回数を可視化する発表支援システムを開発した [1]．先行システムにより，発表者が，自ら癖を自覚することを支援できた．しかし，これは発表者に癖が出ていることは示せても癖を矯正することには至ってない．本研究では，さらに，模範となるプレゼン映像を提示することで，プレゼンにおける癖が改善されると考えた．具体的には，発表練習の映像から，発表者の癖を除去した映像を，模範として提示する．先行システムの癖検出機構は，模範映像の生成のための癖の除去に関して，改善の余地がある．本稿では，癖の検出方法の改善および癖を除去した模範映像の生成について示す．

Developing a Training System based on a Habit Less Presentation Movie Generator

Tatsuo Matsuura, Tadachika Ozono, Toramatsu Shintani, 名古屋工業大学大学院情報工学専攻, Dept. of Computer Science, Graduate School of Engineering, Nagoya Institute of Technology.

本稿では，2 章で，発表練習における癖矯正について述べる．3 章で，過去のプレゼン映像から発表者の癖を検出し，癖の除去を行う癖除去機構について述べる．4 章で，癖除去機構を用いた発表練習支援システムの実行例および癖除去機構の評価実験および考察を述べる．最後に，5 章では，本稿をまとめる．

2 発表練習における癖矯正

2.1 プレゼンにおける癖と矯正

本節ではプレゼンにおける発表者の癖について述べる．学会発表などで他の人と自分の知識や成果を共有するために人前で発表を行う機会は多い．多くの発表者は，スライドや口頭により自分の伝えたいことが聴講者に伝わりやすいように心がける．しかし，発表者は多くの聴講者の前で発表する際に，緊張や不安から声が小さくなったり，口癖などの癖が出てしまうことがある．発表者の癖により聴講者の集中力を欠いたり，発表内容の理解度の低下を引き起こす可能性がある．そこで，多くの発表者は本番を想定して事前に発表練習を行う．しかし，発表者が練習中に無意識な癖により自身の癖に気づけなかったり，発表中に客観的に自身の発表を評価するのは困難である．

先行研究では，発表者が発表している様子をプレゼン映像として記録し，発表の癖を検出および癖の出ている箇所や回数を可視化する発表練習支援システムを開発した．しかし，先行システムでは，発表者に癖が出ていることを自覚することができても効果的に癖を矯正することには至ってない．さらに，発表練習支援システムにおいて発表者の声量，視線，早口および口癖について検出を行ったが，聴講者には発表者の声の高さによって声が小さく聞こえたり，発表者の話す間合いによって早口に聞こえたりなど癖の検出結果に対して，聴講者が感じる癖と異なる部分も生じた．そこで，本研究では，先行システムの改善を行い，発表者の癖を矯正することに焦点を置く．

2.2 模範となるプレゼン映像による発表練習

本研究では，発表者の過去の発表の様子を映像として記録したプレゼン映像から癖を削除し，発表の模範となるプレゼン映像を生成する癖除去機構を提案する．発表者がより良い発表を行うため模範となる発表を模倣することは有効であると考える．模範となる発表としてテレビなどのアナウンサーや世界的講演会である Ted Talk[†1] などが挙げられる．これらのプレゼンも模範となると考えるが，さらに，発表者自身の発表内容に基づく模範映像がより有益であると考えた．そこで，発表者の過去の発表の様子を記録したプレゼン映像から癖を削除し発表者に特化した模範的なプレゼン映像を生成することで，模倣のしやすさおよび目指すべき発表を示すことで，癖の矯正につながると考える．

2.3 癖の検出

本研究では，発表者の過去のプレゼン映像から発表者の癖を検出する．本研究でプレゼン映像から検出する癖の種類は，声量，視線，早口，言葉の間および口癖である．

声量に関する癖として，声量不足を検出する．発表者の声量が不足していると聴講者は発表内容の理解度が低下してしまうため声量の検出が必要である．また，単純に音圧 (dB) で声量を検出するだけではなく，人間には，同じ音圧でも聞き取りやすいおよび聞き取りにくい周波数があるため，発表者の声高に合わせて補正を行い聴講者に聞き取りやすい声量かどうか検出を行う．

視線に関する癖として，PC 画面やスクリーンの注視を検出する．発表者の視線が聴講者に向かず，PC の画面やスクリーンの方ばかり視線を向けていると発表の印象が悪い．本研究では，PC の Web カメラから顔認識を行い視線を検出し PC 画面を向いているかどうかを検出する．さらに，Kinect センサーを用いて発表者の顔の向きを検出し，聴講者の方向に向いているかどうかを検出する．

早口の検出では，一定の時間長を持つ音の文節単位であるモーラ (mora) を基準に判定を行う．日本語の場合，仮名一つの単位が 1 モーラに当たり，小書き仮名は 1 モーラとは数えない．例えば,「しゃ」など仮名＋小書き仮名で 1 モーラと数える．本研究では，

†1 https://www.ted.com

プレゼン映像の音声において 1 秒当たりのモーラを算出し，早口の判定を行う．

言葉の間の検出では，発話の区切りの長さを検出する．1 秒あたりのモーラが早口の基準に満たしていなくても，言葉の間が詰まっていると発表内容の理解に悪影響を与える．言葉の間の長さや位置について須藤ら [2] によると，0.4 秒未満の短い言葉の間や発話の区切りとして不自然な位置で言葉の間がとられていたりすると聴講者に発話内容の理解に悪影響を与える．さらに，接続詞の後，名詞の列挙における体言止めおよび名詞＋助詞の後に言葉の間が挿入されることが自然な位置で取られる言葉の間であるとされる．

口癖として「えーと」や「あのー」などの有声休止の検出を行う．口癖が頻繁に発生している発表では，聴講者の集中力を欠いてしまう恐れがある．検出において有声休止の音素と類似している他の単語も存在するため音声認識のみの判定では誤検出を行う可能性が高い．そこで，本研究では，基本周波数の変動が少なくスペクトル包絡の変形が小さい有声休止の特徴量を検出し，見本となる有声休止を抽出する．そして，見本となる有声休止と類似している音声を探索し，有声休止を検出する．

2.4 癖の除去

模範なプレゼン映像を生成するために検出された癖の除去を行う．本研究では，プレゼン映像中の音声情報に対して癖の除去を行う．発表者は，癖が除去された音声情報を模倣しながら発表練習を行うことで発表の質の向上を図る．声量に対しては，声量が不足している部分を視覚的に指摘を行う．早口に対しては，最初に 1 秒当たりのモーラによるプレゼン映像の発話速度の調整を行う．本研究では，1 秒当たりの基準のモーラ数を 5 モーラとして考え発話速度の調整を行う．しかし，発話速度の調整をして音声の引き伸ばしを行うことで本来の発表者の声の声高と変わってしまうため，声高を変えずに発話速度の調整を行う．

そして，言葉の間は音声認識結果に対して，言葉の間が挿入されるべき位置に適切な長さの言葉の間がない場合は，言葉の間の調整を行う．口癖に対しては，口癖が検出された位置の音声情報を削除することで癖の除去を行う．発表者は癖が除去された音声を聴きながら発表練習を行い，癖の矯正を図る．

表 1　A 特性補正値

周波数 (Hz)	補正値 (dB)
125	-16.1
160	-13.4
200	-10.9
250	-8.6
315	-6.6
400	-4.8
500	-3.2
630	-1.9
800	-0.8
1000	0.0

3 癖除去機構

本節では，過去のプレゼン映像から発表者の癖を検出し，癖の除去を行う癖除去機構について述べる．癖除去機構では，過去のプレゼン映像から声量，視線，早口，言葉の間および口癖の検出を行う．そして，本研究では，音声情報に対して癖の除去を行う．発表者の癖の検出後，検出された癖に対して修正および削除を行う．

3.1 音声区間の取得

プレゼン映像の音声情報に対して，癖の検出を行う前に音声区間の検出を行う．本研究では，大語彙連続音声認識エンジンである Julius[†2] を用いて音声区間の検出を行う．Julius では，音声信号の振幅と零交差数に基づいて音声の開始位置と終了位置を検出する．さらに，振幅と零交差によって検出された音声情報に対してガウス混合分布モデル (GMM : Gaussian mixture model) を用いて音声区間の判定を行う．音声 GMM と非音声 GMM から，短時間フレームごとに特徴量を抽出する．各 GMM の尤度計算を行い，音声 GMM と非音声 GMM の尤度比から音声区間の

†2　http://julius.sourceforge.jp

開始・終了を判定し，音声区間を取得する．

3.2 声量の検出および警告

本研究では，発表者の声量を音の音圧である dB(デシベル) 単位で発表者の声量を検出する．一つの音声区間に対して dB の平均を算出する．周波数と音圧のレベルを等高線で結んだ等ラウドネス曲線 (国際規格:ISO 226:2003) に基づいて声量の検出を行う．癖除去機構では等ラウドネス曲線を考慮した A 特性音圧レベルの平均で声量の計測する．これは，発表者の声の高さに個人差があり，人間は周波数によって感じる音の大きさが異なるからである．表 1 は，A 特性音圧レベルにおいて男声と女声の周波数帯と考えられる 125Hz から 1,000Hz の A 特性補正値を示す．計測された声量に対して A 特性補正値を与え補正声量で判定を行う．本研究では，補正声量の閾値を 70dB とする．図 1 より，125Hz から 1,000Hz の範囲において A 特性補正値を最小二乗法により A 特性補正値 c_i を i 番目の音声区間において，周波数の平均 h_i として，式 (1) とする．さらに，i 番目の音声区間において，v_i を声量，c_i を A 特性補正値として補正声量 V_i を式 (2) で示す．補正声量が閾値を超えていない場合は発表者に対して警告を行う．

$$c_i = 1.7 \times 10^{-2} h_i - 1.4 \times 10 \tag{1}$$

$$V_i = v_i - c_i \tag{2}$$

3.3 視線の検出および警告

視線検出は，プレゼン映像から顔認識を行い，発表者の目の位置を検出することで PC 画面に視線が向いているかどうかを検出する．Jason ら [3] の Constrained Local Model(CLM) を用いた顔特徴追跡を用いて検出を行う．顔のサンプルデータから主成分分析によって基底を求めており，顔の特徴点から目の座標から目線を取得する．顔の特徴点は 66 点あり，$X = (x_0 \cdots x_{65}\ y_0 \cdots y_{65})^{\mathrm{T}}$ より，目の座標を目の周りの座標の重心から求める．(式 (3)，式 (4))

$$(x_r, y_r) = \frac{1}{6} \sum_{i=36}^{41} (x_i, y_i) \tag{3}$$

$$(x_l, y_l) = \frac{1}{6} \sum_{i=47}^{42} (x_i, y_i) \tag{4}$$

目の座標が Web カメラ範囲内にある場合は，PC の画面を見ていると判定を行う．さらに，発表者が聴講者の方に向いているかどうか判定を行う．Microsoft Xbox One Kinect センサーを発表者の前方に置き骨格情報から顔の向きを検出する．Kinect センサーにおいて Face Tracking SDK を用いることで顔の向きを取得し聴講者に向いているかどうか判定を行う．視線が聴講者に長時間向いていない場合は，PC 画面およびプロジェクタ画面上に警告を表示する．

3.4 言葉の間の検出および修正

言葉の間の自動調整を行うためにプレゼン映像の音声情報に対して音素情報の取得を行う．最初に，Julius の単語・音素セグメンテーションキット †3 を用いて音声情報に対して音素のラベリングを行う．次に，ラベリングされた音声から音素情報を取得する．その後，Julius で得られる音素情報から講義映像の発話内容の品詞を解析し，接続詞の後，名詞の列挙による体言止めおよび名詞＋助詞の後に言葉の間が取られているか判定を行う．さらに音素情報から音素の時間区間を算出し，言葉の間が 0.5 秒取られているかを判定し，0.5 秒未満の場合は，0.5 秒になるまで無音を追加する．最後に，言葉の間が取るべきところではない言葉の間がある場合は，言葉の間の削除を行う．

3.5 早口の検出および修正

早口の検出として，プレゼン映像の音声区間情報および音素情報から 1 秒当たりのモーラを算出し，基準のモーラに対して発話速度の変更割合を算出する．

$$R_i = \frac{t_i}{p_i} m \tag{5}$$

†3 http://julius.osdn.jp/index.php?q=ouyoukit.html

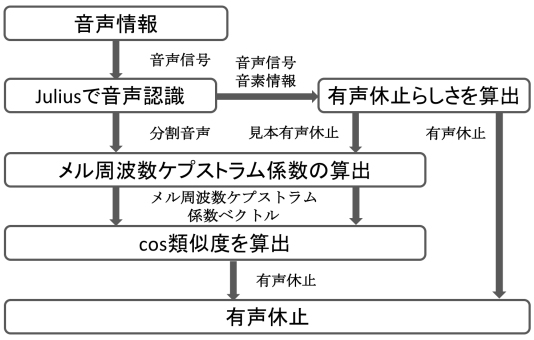

図 1　有声休止の検出

i 番目の音声区間において，発話速度の変更割合の R_i をここで，t_i は音声区間時間，p_i は音素数，m はモーラ基準数である (式 (5))．本研究では，m = 5 mora/s を聞き取りやすい基準として早口の検出および修正を行う．早口の調整手法として発話速度の変更割合の R_i を基準にピッチ同期重畳加算 (PSOLA : Pitch Synchronous Overlap and Add) [4] を用いて，音声情報の声高を変えずに音の伸縮を行う．PSOLA は，波形の零交差点であるピッチマーク毎にハミング窓で波形を切り出し，ピッチ周期とフレーム時間長に合わせて切り出した波形を重畳加算し，合成波形を生成する．

3.6　口癖の検出および削除

口癖の検出の処理の流れを図 1 に示す．最初に，プレゼン映像の 1 つの音声区間および音素情報から有声休止になる可能性の高い音声区間を抽出する．有声休止は,「えーと」や「あのー」など，一つの音声区間に対して音素数が少ないため，1 つの音声区間に 5 つ以下の音素を含む音声区間を抽出する．そして，有声休止を検出するため後藤ら [5] の有声休止の検出手法を用いる．有声休止を基本周波数の変動が少なくスペクトル包絡の変形が小さい特徴から検出を行う．基本周波数の変動量 $A_s(t)$ とスペクトル包絡の変形量 $A_s(t)$ の短時間平均 $S_f(t)$，$S_s(t)$ を算出する．

$$P_{f_p}(t) = exp(-\frac{(RS_f(t) + (1-R)S_s(t))^2}{W^2}) \quad (6)$$

有声休止らしさである $P_{f_p}(t)$ が高い値を取り続けた音声区間が有声休止を含んでいる可能性が高い (式 (6))．そこで，有声休止らしさの高い音声区間を見本の有声休止としてプレゼン映像の全体の音声情報に対して，音声の類似度を算出し有声休止を検出する．音声の類似度は，声道特性の特徴量であるメル周波数ケプストラム係数 (Mel-Frequency Cepstrum Coefficients) [6] から求める．見本の有声休止と音声情報を分割した分割音声からメル周波数ケプストラム係数ベクトルを算出する．算出されたメル周波数ケプストラム係数ベクトルが 20 次元のベクトルがフレーム数分あり，膨大なベクトル数になるため，k-means でベクトル分類する．最後に，メル周波数ケプストラム係数ベクトルに対して Cos 類似度を算出する．見本有声休止との Cos 類似度から有声休止を検出する．見本の有声休止のメル周波数ケプストラム係数ベクトルを M_s，i 番目の音声区間においての音声情報のメル周波数ケプストラム係数ベクトルを M_i として，類似度を S_i を式 (7) で示す．そして，検出された口癖の位置の音声情報を削除し，口癖の削除を行う．

$$S_i = \frac{M_s \cdot M_i}{||M_s||||M_i||} \quad (7)$$

4　癖検出機構の実行例と評価実験

4.1　プレゼン映像の閲覧とリアルタイム検出

図 2，図 3 および図 4 を用いて本システムの実行例を説明する．図 2 では，発表者が過去のプレゼン映像を閲覧している様子を示している．図 2(A) では，発表者が用いた発表のスライドが表示されている．スライドの音声情報は連動して表示されている．図 2(B) では，音声の波形と連動して Julius の音声認識結果が表示されている．図 2(C) では，音声の波形が表示されている．発表者は，波形を選択することで途中から音声の再生を行えたり，検出して欲しい有声休止を指定することができる．図 2(D) では，声量，早口，視線および言葉の間の癖が出ている回数と範囲

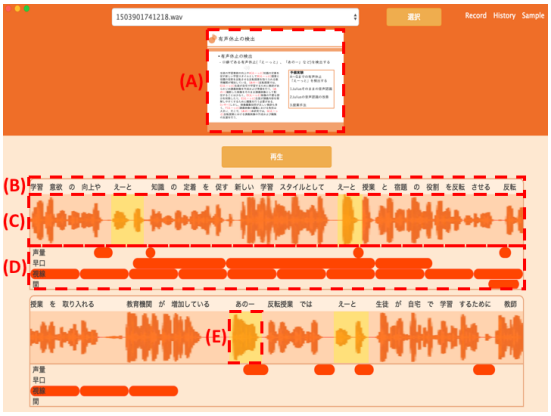

図 2　過去のプレゼン映像の閲覧

図 4　発表練習の様子

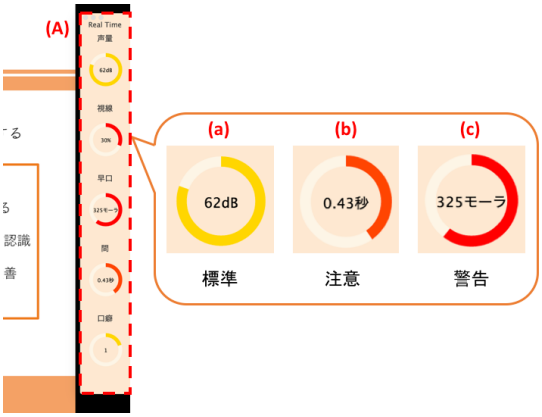

図 3　リアルタイムに発表者の癖を検出

が波形の位置と連動して表示されている．図 2(E) では，見本の有声休止から口癖が出ている波形の部分を検出し，表示されている．発表者は，過去のプレゼン映像とともに可視化された癖の箇所や回数を見直し，効果的に発表練習を行うことができる．また，図 2 で検出されている癖の除去を行った音声を出力し，発表練習で活用することができる．

図 3 および図 4 では，リアルタイムに発表者の癖を検出している様子を示している．本システムでは，リアルタイムに発表者の癖を 5 秒ごとに検出し，発表者の癖の検出状況をスライド表示の空き領域 (図 3(A)) に表示する．癖が検出されているレベルとして標準 (図 3(a))，注意 (図 3(b)) および警告 (図 3(c)) で表示する．また，本システムでは，発表者の声量，視線，早口，言葉の間および口癖をリアルタイムに検出している．図 4 では，実際に発表練習を行なっている様子を示している．発表者に前方に Kinect を設置し，癖のリアルタイム表示を見ながら発表練習を行うことができる．発表者は本システムによりリアルタイムに検出された癖の状況を確認しながら効果的に発表練習を行うことができる．

4. 2　評価実験

癖除去機構の癖検出精度について評価実験を行う．本実験では，癖除去機構において，音声情報を対象に早口，声量および口癖についての癖の検出精度について評価を行う．被験者は，20 代男性 5 人で，初めて見るスライドを用いて PC の前で発表を行う．早口および声量については聴講者が 10 秒ごとに癖が出ている部分を判定し，正解データとして扱い，癖除去機構の検出精度を評価する．また，口癖については，有声休止である「えーと」,「あのー」,「いやー」および「そのー」について検出できているか評価する．

被験者 5 人には，初めて見るスライド 3 枚分をプレゼン映像として記録を行い，癖検出機構によって発表者の癖を検出した．音声データとしては，合計 14 分 54 秒，181 区間，口癖は 120 箇所含まれている．早口および声量については，正確度 A，口癖については，適合率 P と再現率 R を評価する．それぞれの式を式 (8)，式 (9)，式 (10) で示す．

$$A = \frac{\text{正しく検出した区間の数}}{\text{区間の数}} \quad (8)$$

$$P = \frac{\text{正しく検出した有声休止の数}}{\text{有声休止として検出した数}} \quad (9)$$

$$R = \frac{\text{正しく検出した有声休止の数}}{\text{有声休止の総数}} \quad (10)$$

実験の結果，早口の正確度は，70.7 % (128 ／ 181)，声量の正確度は，76.8 % (139 ／ 181)，口癖の適合率は，81.3 % (61 ／ 75)，再現率は，50.8 % (61 ／ 120) であった．

4.3 考察

本システムの実行例および評価実験について考察する．本システムの実行例では，発表者の発表の中で癖が出ている部分を検出し可視化することができた．発表者は癖が出ているタイミングや傾向を確認し，癖に注意しながら効果的に発表練習を行うことができる．さらに，リアルタイムに癖を検出し発表練習中でも客観的に自身の発表の評価を行うことができる．

また，評価実験より発表中の発表者の癖を検出することができた．声量においては，音圧 (dB) のみの検出ではなく A 特性補正値を与え，人間の聴覚に考慮した声量検出になっており，音圧が小さくても声がよく通る周波数のため声を小さく感じない音声に対しても適切に判定することができた．早口において単純に秒単位のモーラ数のみならず，言葉の間を考慮することで，言葉の間が詰まっていることで感じる早口を検出することができた．しかし，声量および早口において，聴講者が声が小さいおよび早口であると感じるタイミングとずれている部分があるためシステムの調整が必要であると感じた．口癖においては，基本周波数の変動が少なくスペクトル包絡の変形が小さい有声休止の特徴から見本の有声休止を抽出し，メル周波数ケプストラムによる人間の声道を考慮した特徴量から類似度を求め判定することができた．しかし，中には早口の部分で言葉の間がなく，有声休止の前後の単語に影響され検出することができなかった有声休止も存在した．

以上により過去のプレゼン映像から癖検出を行い，癖が削除された模範的なプレゼン映像を生成する環境を整えることができた．今後，癖を除去した模範的なプレゼン映像での効果的な発表練習について研究および評価を行いたい．

5 おわりに

本稿では，発表者の発表中の癖を矯正するため，過去のプレゼン映像から癖を削除したプレゼン映像を生成する癖除去機構について提案した．本研究では，癖検出機構によりプレゼン映像から視線，声量，早口，言葉の間および口癖の検出を行い，音声情報に対して，検出された癖を修正および削除を行った．発表者の過去のプレゼン映像から癖を除去した発表者に特化した模範的なプレゼン映像を生成することで癖の矯正が期待される．

謝辞 本研究の一部は JSPS 科研費 JP15K00422, JP16K00420 の助成を受けたものです．

参 考 文 献

[1] 松浦辰雄, 大囿忠親, 新谷虎松, “スライド発表における癖検出機構に基づく発表練習支援システム”, 第 79 回全国大会講演論文集, 2017 (to appear)

[2] 横井聖宏, 馬場康輔, 須藤秀紹, 山路奈保子, “発話中の「間」がプレゼンテーションに対する聴衆の支持に与える影響 書評ゲーム『ビブリオバトル』の発生音声録音データ分析による考察”, 日本感性工学会論文誌, Vol.15, No.3, pp.363-368, 2016.

[3] Jason M. Saragih, Simon Lucey, Jeffrey F. Cohn, “Deformable Model Fitting by Regularized Landmark Mean-Shift.”, International Journal of Computer Vision, Volume 91, pp.200-215, 2011.

[4] Eric Moulines ,Francis Charpentier, “Pitch-ssynchronous waveform processing techniques for text-to-speech synthesis using diphones.”, In Speech Communication 9, pp.453-467, 1990.

[5] 後藤真孝, 伊藤克亘, 速水悟, “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, Volume 83, No.11, pp.2330-2340, 2000.

[6] Md.Sahidullah, Goutam Saha, “Design, analysis and experimental evaluation of block based transformation in mfcc computation for speaker recognition.”, In Speech Communication 54, pp.543-565, 2012.