社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

ボトムアップクラスタリングを用いたシャドーイング音声の自動評定

下村　直也† 峯松　信明† 山内　豊†† 広瀬　啓吉†††

† 東京大学大学院新領域創成科学研究科 〒 277–8561 千葉県柏市柏の葉 5–1–5
†† 東京国際大学商学部 〒 350-1197 埼玉県川越市的場北 1-13-1
††† 東京大学大学院情報理工学系研究科 〒 113–0033 東京都文京区本郷 7–3–1
E-mail: †{shimo,mine,hirose}@gavo.t.u-tokyo.ac.jp, ††yyama@tiu.ac.jp

あらまし　コミュニケーション力を重視する昨今の外国語教育において，シャドーイングという学習方法が広がりを見せている．シャドーイングとは，外国語音声を聞きながらほぼ同時にその発話を繰り返して発声することで，発音能力と聴取能力とを同時に鍛える訓練方法である．提示音声の発声速度に追従する必要があるため，非常に負荷の高いタスクであり，特に初学者の場合，シャドーイング音声は非常に崩れた／不明瞭な音声となる．外国語発音の自動評定を行なう場合，母語話者の読み上げ音声から構築された HMM を用いて行なうことが多い．しかし，1) 発話スタイルが読み上げ調と大きく異なる，2) 任意の外国語のシャドーイング音声評定を行なうためには，提示音声及びシャドーイング音声のみを用いて評定を行なえることが望ましい，といった点を考慮し，HMM の事前学習を必要としないシャドーイング音声自動評定手法を提案する．ここでは，音響分析のみに基づくボトムアップクラスタリングを行ない，発声中に観測されるセグメント数に着眼する．実験の結果，自動評定スコアと手動評定スコア（及び TOEIC スコア）との間に比較的良好な相関が観測された．また，作成したシャドーイング学習支援システムについても報告する．
キーワード　シャドーイング，ボトムアップクラスタリング，自動評定，調音努力，外国語学習支援

Automatic scoring of language learners' utterances generated through shadowing based on bottom-up clustering

Naoya SHIMOMURA†, Nobuaki MINEMATSU†, Yutaka YAMAUCHI††, and Keikichi HIROSE†††

† Grad. School of Frontier Sciences, Univ. of Tokyo, 5–1–5, Kashiwanoha Kashiwa, Chiba, 277–8561 Japan
†† Faculty of commerce,Tokyo International Univ., 1–13–1, Matobakita Kawagoe, Saitama, 350–1197 Japan
††† Grad. School of Info. Sci. and Tech., Univ. of Tokyo, 7–3–1, Hongo, Bunkyo-ku, Tokyo 113–0033 Japan
E-mail: †{shimo,mine,hirose}@gavo.t.u-tokyo.ac.jp, ††yyama@tiu.ac.jp

Abstract　Recently, shadowing has attracted much attention in the field of teaching and learning foreign languages. Shadowing is a kind of "repeat-after-me" type exercise but it requires learners to repeat each word before hearing its final syllable. Shadowing is thought to raise both pronunciation and hearing abilities effectively. Since learners have to follow the speaking rate of an input native utterance, especially in the case of beginners, their pronunciation often becomes inarticulate and corrupt. To build an automatic scoring system of non-native utterances, HMMs which have been trained with native "read" speech, are often used. In this study, considering two facts of shadowing; 1) the speaking style of learners in shadowing is very different from that of "read" speech, 2) it is desirable to build a scoring system which requires only an utterance pair; a native utterance presented to learners and a learner's utterance generated in response to the native utterance, a new method is proposed here for automatic scoring of utterances in shadowing. The new method does not use any acoustic models such as HMMs at all and just compares the two utterances based on time-constrained bottom-up clustering. Experiments show that rather good correlation is found between automatic scores and manually-rated scores (or TOEIC overall proficiency scores).
Key words　shadowing, bottom-up clustering, automatic scoring, articulatory effort, support technology for teaching and learning foreign languages

1. はじめに

近年，言語教育においてシャドーイングが注目されている．シャドーイングとは，聴取した（母語話者により発声された）外国語音声を即座に繰り返して発声する外国語聴取・発音訓練法である．元来，同時通訳者の訓練として広く行なわれていたが（この場合，故意に delay を置いて通訳するなど，認知的により高いタスクを要求する），外国語学習においてもシャドーイング学習の効果が認められるようになった [1], [2].

学習初期段階の日本人が英語を発声すると，カタカナ英語と呼ばれる発音となることがある．認知心理学的には「英単語の発音が，日本語の音韻に変換された状態で，長期記憶中の心的辞書（メンタルレキシコン）に保持されていることに起因する」と考えられている．シャドーイングは，心的辞書から語彙情報を検索する時間を十分に与えずに発声を要求するため，入力音声の音的イメージをそのまま再生させることに繋がり，母語の音韻体系に引きずられることなくスピーキング能力を向上させることができる，と考えられている [1], [2].

さらに，シャドーイングはリスニング能力の向上ももたらす．リスニングは「知覚」と「理解」から構成されているが，両段階において，認知資源を消費する．シャドーイングは，母語話者の発音を繰り返して聞くことで音声知覚過程を鍛え／自動化し，同時に，スピーキングを通して正確な発音（音的イメージ）を心的辞書に定着させることで，理解の段階により多くの認知資源を割り当てられるようになる．これらの結果，リスニング能力についても，その向上が期待できる [1], [2].

このようにシャドーイングは，スピーキング／リスニング能力を同時に訓練できるため，コミュニケーション能力を重視する近年の外国語学習において広がりを見せている．学習意欲維持のためには学習者が自らの習熟度を把握し，また教師側は，学習者発声を短時間で評定し教示する必要がある．しかし，シャドーイングは非常に負荷の高い訓練法であり，シャドーイング音声は一般にかなり「崩れた」音声となる．人手でこれらを逐一評定することは膨大な時間を要するため，発音評定技術を用いた自動化が望まれるところである．しかしシャドーイング音声は，従来の評定技術が対象としてきた比較的「奇麗な」音声とはかなり異なる．筆者らの知る限り，シャドーイング音声を対象とした自動評定手法は提案されていない．

筆者らは，笑い声や鳴き声といった「崩れた」音声をも対象としうる技術として，ボトムアップクラスタリングに基づく教師無しセグメンテーションを検討してきた [3], [6]．本稿では，この技術をシャドーイング音声の自動評定に応用する．その結果，自動評定スコアと教師による手動評定スコアとの相関は 0.79 と良好な値が得られることを示す．また，話者毎の自動評定スコアと TOEIC スコアとの相関も，シャドーイング対象の文を適切に選定することで相関 0.75 が得られ，TOEIC スコアの簡易自動推定手法としての可能性についても論じる．最後に，これらの自動評定技術を組み込み，外国語学習支援システムとして構築したアプリケーションについて簡単に紹介する．

2. シャドーイングの学習レベル及び，評定方法

2.1 学習者レベルに応じて異なる形態を示すシャドーイング

初期段階の英語学習者のシャドーイング音声は，非常に崩れた発声となる．耳から入る音声の知覚が十分に自動化されておらず，更には，英語の各音韻・音節を生成するための調音運動が十分に習得できていないため，時に言い黙り，言い淀みが生じる．逆に，熟練者のシャドーイング音声は調音努力が十分になされ，流暢／明瞭な発声となる．

同じシャドーイングであっても，学習者レベルに応じて下記のような段階的な発声スタイルが観測，定義されている．

（1） マンブリング

声に出すか出さないかの小声で，口をもぐもぐ動かす発声．内容の伝達は非常に不完全となる。低レベル話者の発声に現れる．

（2） プロソディック・シャドーイング

意味はともかく，イントネーションやリズムを追うように発声するシャドーイング．日本語的な発声にならぬよう,「英語らしさ」を重視して発音を行なう段階である．

（3） コンテンツ・シャドーイング

韻律調整が正しく行なわれつつ，文の意味内容も捉えながら行なうシャドーイング．時として，ポーズが多くなり，句毎に単語群を一度に発声するため話速が早くなることがある．

2.2 手動評定手法

上述のような発声スタイルの違いを考慮しつつ，シャドーイング音声評定方法が玉井によって 2 種類提案されている（音節法及びチェックポイント法 [2]）．また，これらの手法の問題点を考察し，3 つ目の手動評定方法として「全単語法」を考える．

2.2.1 音節法

音節法とは，英語における発話の最小単位と考えられている音節毎の正誤を判定する方法である．素材となる外国語テキストの書き起こしをもとに，1 音節語はそのままに，2 音節以上の単語は各音節毎に分け，評定を行なう．評定単位が単語より小さく，評定の信頼性が保たれる．しかし，採点者は音声を音節毎に区分化，評定する必要があり，時間的コストや体力的負担を被ることとなる．そのため，必ずしも実用性が高い方法ではないと筆者らは考える．

2.2.2 チェックポイント法

音節法と違って，単語毎に評定する簡便法として，チェックポイント法が提案されている．英語テキストの全単語を n 単語毎に，その単語が正しく発声できているかどうかを判定する．[2] では，全単語が 350 語以上で，各文が 8 単語程度の長さで構成されている場合に n=5 を採用している．この場合，音節法との評定結果の相関として 0.89 が示されている [2].

2.2.3 全単語法

音節法は一見精度が高そうに見えるが，because を become とシャドーイングした場合，音節法では 50 %の正答を与えることになり，チェックポイント法では 0 %になる．because を become とシャドーイングするのは，明らかに単語を取り違えており，英語をコミュニケーション・ツールとして捉え，実践的コミュニケーション能力の養成を目的とする，近年の英語教育

の方向性と乖離している．また，n の設定方法は十分に明らかとなっていない．これらを考慮し，本稿で行なう手動の評定は n=1，即ち全単語に対して「その単語が発声できているか」を判定し，手動の評定スコアとした．この場合，その単語として意図されたと思われる発声であれば，正解として判定している．

3. 時間制約付きボトムクラスタリングによる教師無しセグメンテーション

3.1 時間制約付きボトムアップクラスタリング

教師無しセグメンテーションの先攻研究の多くは，局所的なスペクトル変化の大きい時点をセグメント境界とする方法が多い [4], [5]．これに対して筆者らはスペクトル的に類似し，かつ，隣接して存在するフレームをマージする形で纏め上げ（ボトムアップクラスタリング），音声ストリームの大局的な階層構造を抽出する形で教師無しセグメンテーションを実装した [3]．更にマージ対象となる 2 セグメント（クラスタ）の選択基準として種々の統計量に基づく尺度を検討し [6]，提案手法が従来の局所的なスペクトル変化に基づく手法よりも優位であることを示した [6]．なお，本稿では実装が容易な Ward 法によるボトムアップクラスタリングに基づく手法 [3] を採用する．

Ward 法はユークリッド距離に基づくクラスタリングであり，2 つのセグメントをマージした際の「群内偏差平方和の増加量」を両者の非類似度と定義している [7]．この非類似度が最も低い（即ち，最も類似している）セグメント同士をマージさせ，最終的には 1 つのセグメントへと纏め上げる．今，時間的に隣り合うセグメント p，$p+1$ をマージして新しいセグメント $r(=p \cup p+1)$ を作ることを考える．p の偏差平方和を $E(p)$ とおくと，r を生成した際の偏差平方和の増分 $\Delta E(p, p+1)$ は

$$\Delta E(p, p+1) = E(r) - \{E(p) + E(p+1)\} \qquad (1)$$

となる．各段階でマージによる偏差平方和の増分 $\Delta E(p, p+1)$ が最小となる p と $p+1$ をマージする．以上の操作は，フレームの部分系列をより大きな纏まりとして捉えることに相当する．

3.2 クラスタリングの停止条件

ある段階において，各セグメントが凡そ各音素に対応している状態を考える．この場合，次のマージ操作は異なる音素を強引にマージすることを意味する．ある話者が生成する各音素の音響特徴を考える．この場合，任意の 2 音素間距離（重心間距離）の最小値は，凡そ話者非依存であると仮定する．その結果，どの話者の発声した音声であっても，音素に対応した形でクラスタリングされた状態に対する次のマージ操作は，比較的大きな更新コスト（群内偏差平方和の増加量）を呈するはずである（$E(p \cup p+1) \gg E(p) + E(p+1)$）．

そこで，$\Delta E(p, p+1)$ に対応する閾値 K を定め，各セグメントが凡そ音素に対応したときに自動的に停止する方法を検討した [3]．具体的な自動セグメント例を図 1 に示す．

図 2 は，提示音声及び，その音声を TOEIC421 点，202 点の英語学習者がシャドーイングした発声をボトムアップクラスタリングした結果である．音声横軸にクラスタ数，縦軸に Ward 法の更新コスト，つまり $\Delta E(p, p+1)$ をプロットした．閾値

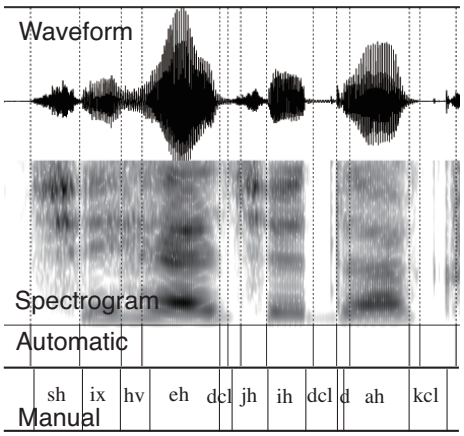

図 1　自動セグメンテーションの一例

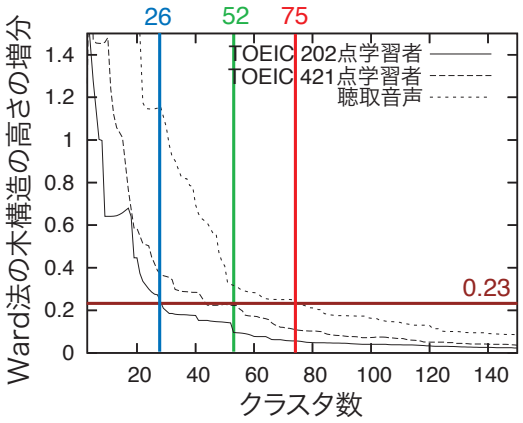

図 2　シャドーイング音声の分析結果

$K = 0.23$ でクラスタリングを停止させた時のクラスタ数を見てみると，提示音声が 75 クラスタ，TOEIC421 点話者の発声が 52 クラスタ，TOEIC202 点話者の発声は 26 クラスタとなっている．同一文を発声した場合でも，個々の音を明瞭に区別（調音）して発声した場合，停止時のクラスタ数は多く，個々の音が不明瞭に発声されれば，停止時のクラスタ数は少なくなる．

3.3 音響事象群における事象間距離と調音努力

「個々の音が音響的に明瞭に区別できていない」場合，それは調音的にも区別できていないことを意味する．音響事象群に対して，全ての二事象間距離を求めて幾何学構造（距離行列）として事象群を表象する音声の構造的表象が提案されている．この場合，距離行列から構造のサイズ（構造の半径に相当する）が求まるが，この量が，凡そ調音努力に相当する定量的尺度になることが実験的に示されている [8]．調音努力とは，個々の音を区別して調音するために行なうべき調音運動量と解釈される量である．例えば母音構造を考えれば，その中心には弱母音，即ち，最も脱力した状態で発声される母音が位置しており，その他の母音は，その母音を発声すべく調音努力を払って声道形状を制御して生まれる音である．事象間距離に対する考察は，読み上げ音声／話し言葉音声の間でも行なわれており，当然話し言葉の方が「なまけ」などの理由で事象間距離が小さくなる [9]．これらを考慮すると，事象間距離の大小を通して，発声時に払われた調音努力の大小を推定することは十分妥当である．

結局，適切な固定閾値の下でクラスタリングを停止させ，そ

表 1　被験者の TOEIC スコア

熟練度別学習者数	素点	平均点
上位者 4 名	968, 955, 940, 895	940
中位者 3 名	432, 427, 421	427
下位者 4 名	301,252, 202, 197	238

表 2　音響分析条件

サンプリング	16bit / 16kHz
窓及び窓長	ハミング窓 / 16msec
シフト長	10msec
音響パラメータ	MCEP 1〜12 次元
クラスタリング停止条件	閾値 $K = 0.23$

の時のセグメント数の大小を議論することは，その発声において払われた調音努力の大小を推定することに相当する．言い換えれば，与えられた発声に対して，どの程度「滑舌の良い」「呂律の回った」発声であったのかを推定することになる．シャドーイング音声は，習熟度が低ければ「もごもごした」音声であり，高ければ「はきはきした」音声となることを考えると，筆者らが提案する教師無しセグメンテーションはシャドーイング音声の評定に非常に相性の良い技術であると言える(注1)．

4. シャドーイング音声の自動評定実験

4.1 シャドーイング音声の収録と手動による評定

日本人学習者 11 名にシャドーイングを行なわせた．TOEIC テスト（990 点満点）における上位 4 名，中位者 3 名，下位者 4 名を対象とした．彼らの TOEIC スコアを表 1 に示す．シャドーイングは初めて，或は，ほぼ未経験の被験者を採択した．

シャドーイング用に提示した音声は，1 名の男性米語母語話者が読み上げた音声であり，全 10 文（174 単語）からなる平易な速読用の英文である．日本の魚に関するトピックで内容的に親密度は高いが，全ての被験者にとって初見である．平均話速は 140 語/分であった．11 名による合計 110 発話のシャドーイング音声を実験に用いた．収録で用いた教室では空調等の定常雑音が随時発生しており，提示音声の収録環境とは異なる．

手動による評定作業は，小・中・高校，及び，大学で英語授業を実践してきた英語教育の専門家（第三著者）によって全単語法で行なわれた．各発声に対して，その単語として意図された単語発声の個数を数え上げ，その文に含まれる語数で割った値（百分率）を，その発声のスコアとした．ただし，言い直した場合に言い直す前の単語は不要語として −1 語，動詞や複数語尾の誤りは −0.5 語として評定した．1 名分（約 2 分の音声データ）の評定に 1 時間程の時間を要したことを明記しておく．

4.2 音響分析及びクラスタリングの諸条件

各種の分析条件を表 2 にまとめる．スペクトル変化を捉える音響特徴量として，聴覚特性を考慮したメルケプストラムを用いた．なお本稿では，スペクトルの安定区間を纏めることでセグメント数を得ることを検討していること，そして，収録されるシャドーイング音声によってはパワーが大きく異なることから，パワー項（MCEP の 0 次項）は用いていない．

クラスタリングの停止条件である閾値 K は，TIMIT train パートの全 4620 発話に対して，正解音素数と自動推定音素数との相関が 0.83 と最も高かった $K = 0.23$ を用いた．

(注1)：なお，ボトムアップクラスタリングを行なわず，初期の $N \times N$ 距離行列（N= 総フレーム数）における構造サイズをもって，与えられた発声の調音努力は推定可能と考えられるが，本稿では固定閾値におけるセグメント数をもって調音努力と解釈した．

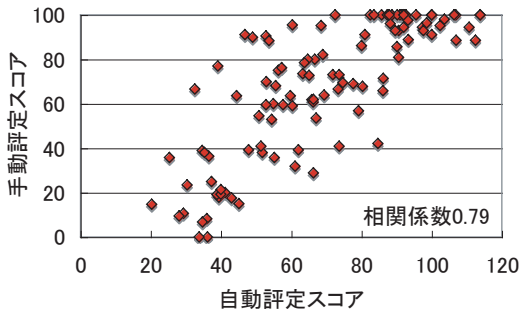

図 3　自動評定スコアと手動評定スコアの関係

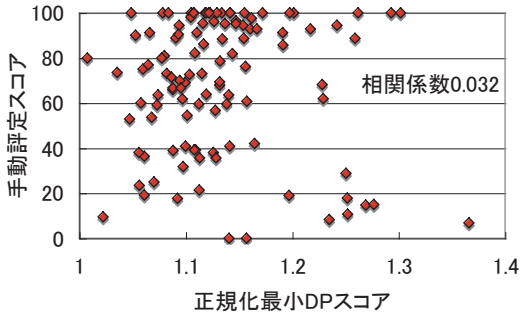

図 4　正規化 DP スコアと手動評定スコア

4.3 各シャドーイング発声の自動評定結果

収録した 110 発声及び，提示した 10 発声を各々クラスタリングし，自動停止した時のセグメント数を算出した．そして「提示音声のセグメント数」に対する「発声者が生成したセグメント数」を百分率で算出し，これをその発声の自動評定スコアとした．自動評定／手動評定スコアの関係を図 3 に示す．全体の相関は 0.79 となり，高い相関を得た．

4.4 DP マッチングによる自動評定結果

提案手法は，提示音声とシャドーイング音声間で一切音響的照合は行なっていない．シャドーイングは提示音声をそのまま真似る（再生する）という側面を有している．そこで比較対象として，提示音声と再生音声を音響的に照合することでスコアを算出する DP マッチングも行なった．この場合，両音声間の DP スコアが小さいほど「音響的に」類似している音声となる．その結果，DP スコアと手動評定スコアとは負の相関が見られるはずである．図 4 に正規化 DP スコアと手動評定スコアの関係を示す．相関係数 0.032 というほぼ無相関の結果を得た．DP マッチングは，話者性の違いによってスコアが大きく変動する．不一致（ミスマッチ）問題が浮き彫りになった．

5. TOEIC スコアの自動評定

本節では，リスニング，スピーキングの複合的学習方法であるシャドーイングの自動評定結果と，総合的英語習得度を反映

熟練度別学習者数	素点	平均点
母語話者 2 名	990, 990	990
上位者 5 名	968, 955, 940, 895, 825	917
中位者 9 名	625, 601, 592, 581, 512	
	436, 432, 427, 421	514
下位者 11 名	395, 367, 308, 301, 289, 278	
	275, 252, 202, 197, 158	275

表 3　被験者の 27 名の TOEIC スコア

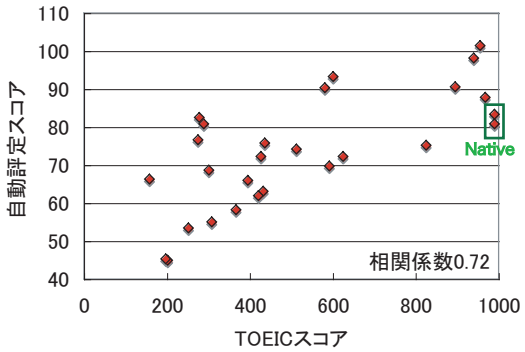

図 5　TOEIC スコアと自動評定スコアの関係

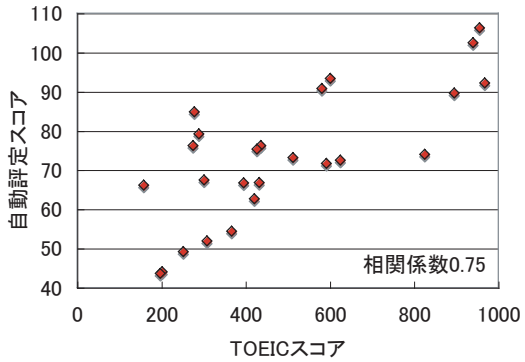

図 6　文選定を行なった上での TOEIC スコアと自動評定スコアの関係

すると言われる TOEIC スコアとの関係について検討する．

ここでは，母語話者 2 名，TOEIC 上位者 5 名，TOEIC 中位者 9 名，及び TOEIC 下位者 11 名のシャドーイング音声を用意した．本実験に用いたシャドーイング音声の発話者の TOEIC スコアを表 3 に示す．母語話者は TOEIC 満点 (990 点) 相当としている．また，各話者の発話文は前節より 11 文増やし 21 文とした．トピックの内容は前節と同様である．

5.1　自動評定スコアと TOEIC スコアの関係

各発話を自動評価に基づきスコア化し，全発話を平均化することで各話者をスコア化した．TOEIC スコアと各話者の得点との相関を図 5 に示す．相関係数は 0.72 である．

5.2　考　　察

5.2.1　母語話者の自動評定スコア低下

TOEIC 上位者の自動評定スコアは上昇し，下位者は下降する様子が分かるが，図 5 を見ると，母語話者の自動評定結果 (四角で囲ってある) が想定される点数より低く算出されている．母語話者は，母語の発話であるため音素間調音結合を滑らかに行なう．これは一種のなまけであり，それが表面化し，点数の低下を招いた可能性が考えられる．また，母語話者の場合，コンテンツ・シャドーイングを行なう傾向があり (ポーズを置いて，一気に句を再生する)，発話速度が上昇する．これも自動評定スコアを減少させることになる．学習者と母語話者で異なるタスクを遂行していると解釈することもできる．母語話者以外の計 25 名による自動評定スコアと TOEIC スコアとの相関係数は 0.74 であった．

5.2.2　発話文の選定

どのレベルの学習者であっても発話可能な発話文 (This is a pen 等)，もしくは困難な文 (専門用語の羅列等) が評価材料となっていた場合，自動評定スコアと TOEIC との相関は低下すると予想される．そこで，21 文中，自動評定スコアの学習者間分散が大きい 12 文 (下位者にとっては難しく，上位者にとっては容易な文) を抽出して自動評定を行なった．母語話者は除いた．文限定時の相関図は図 6 のようになり，相関係数は 0.75 であった．学習話者レベルに応じた文を選定することで，よりコンパクトに自動評定が可能となると考察される．

分散が高くなる発話文の特徴として,「文が短すぎないこと」及び「発話後半に録音される発話であること」を挙げることができる．単語数が数語しかない文の場合，どの学習者のワーキングメモリも不足せず，容易に発話可能となり得点差が発生し難いと考えられる．また，後半の発話文になるほど，学習者の集中力がより問われることになる．熟練度の高い学習者よりも，低い学習者の方がシャドーイングに対する負荷が多くかかると考察できるため，シャドーイングをしばらく行なった後の発声を評価対象とすることは妥当であると思われる．

5.2.3　発話スタイルの違いによる自動評価の変化

図 5 において，TOEIC スコア 300 点程度の学習者の自動評定スコアは 55〜85 点と大きな開きが見られる．50 点台の学習者のシャドーイング音声は，無音区間が多いが，発声された単語は比較的明瞭である．逆に 80 点台のシャドーイング音声は，無音区間は少ないが，ぼやけた発音になっている．50 点台の学習者はプロソディック・シャドーイングもしくはコンテンツ・シャドーイングを行なっており，80 点台の学習者はマンブリングに近い作業を行なっていると言える．本来異なるタスクを同一の枠組みで評価しているため，大きな評価の差を生む結果となった．これを防ぐには，学習者に同じタスクを遂行させるための適切な事前のインストラクションが必要である．また，マンブリング音声に対してクラスタ数が容易に増加しないよう，K を高めに設定する方法も考えられる．今後の検討事項である．

6.　シャドーイング学習支援システム

前述のシャドーイング自動評定技術を組み込んだ，シャドーイング学習支援システムを構築した．本システムは学習者への提示音声と，学習者のシャドーイング音声のみを用いて評定を行なう（不特定話者 HMM などの事前学習を必要としない）ため，そのまま実システムとして提供できる状況にある．本システムのインタフェースの一部を図 7 に掲載する．

本システムでは，様々な学習者のニーズに応えるために下記に列挙する機能を備えている．

- じっくりシャドーイングモード (練習モード)

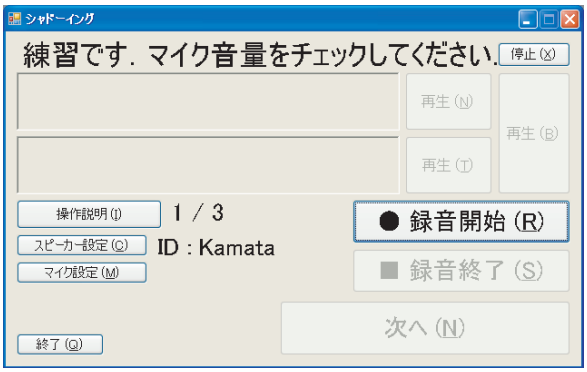

図 7　シャドーイング学習支援システムの UI

提示音声群を選択後，録音開始ボタンによりシャドーイングを行なっていく練習モード．シャドーイング音声は録音され，各文のシャドーイング終了後，提示音声とともに波形が表示される．3 つの再生ボタンを押すことにより，提示音声，シャドーイング音声，もしくはその両方を同時に聞くことが可能であり，学習者は自らの学習の程度を確認できるようになっている．次文の提示開始は学習者のクリックに同期して行なわれる．

- 連続シャドーイングモード (テストモード)

シャドーイングの熟練度を測定するモード．提示音声群を選択すると，その音声群が数秒の間を開けて次々と再生される．学習者のクリック一つで全文に対するシャドーイングが一気に行なわれる．全シャドーイングが終了すると，自動的に本評定システムが稼働し，評定結果を出力する．評定結果は，各発話の得点と予想 TOEIC スコアにより構成される．後者のスコアは，図 5 に対する最小二乗法に基づく線形回帰により予想する．

- 音声再生モード (復習モード)

上記した二つのモードで録音したシャドーイング音声と提示音声の，音声波形の確認と再生が行なえるモード．

- 入力音量及び出力音量調節機能

GUI ベースでスピーカー音量とマイク音量を調節できる．

7. ま と め

筆者らが提案している教師無しセグメンテーション手法が，調音努力（滑舌の良さ）に相当する定量的尺度を提供できることを鑑み，近年注目を集めているシャドーイングに着目し，その自動評定を試みた．その結果，手動評価に凡そ沿った自動評定が可能であることを示した．その一方で，DP マッチングでは自動評定は極めて困難との結果を得た．また，英語力の総合評価を行なう TOEIC スコアとシャドーイング自動評定スコアが良好な相関関係にあることも示した．その上で，構築したシャドーイング学習支援システムについて簡単に紹介した．

本提案手法は言語非依存の技術である．仮に，新たに言語が発見された場合であっても，その直後に，その言語の発音・聴取能力の自動評定が可能となる技術である．また提案手法は不一致（ミスマッチ）問題とは無縁の技術である．距離行列計算で必要なのは，同一話者内での「音と音の距離計算」のみである．DP や HMM のように異なる話者間で「音と音の距離計算」を行なえば，不可避的に不一致問題が発生する．発音評定を行なうシステムを構築する場合，低いスコアを提示された時に，それが学習者の習熟度が低いからなのか，それとも学習者の声質がシステムの学習データに合致しないのか，不明であることがしばしばある．筆者等が提唱する音声の構造的表象を含め，教育応用には，安全かつ健全な技術構築が望まれると考える．

今後の課題をいくつか挙げておく．1) まずは，刺激文レベルの適正化である．学習者群の各レベルへの分類に際して，適した難度の発声文を明らかにしていく必要がある．項目反応理論に基づいて検討することを考えている．2) 異なるシャドーイングタイプに対応できていない可能性がある．評価を行なう際には，マンブリングを行なわないよう指導することが求められる．或は，閾値 K を手動スコア（TOEIC スコア）との相関を最大化するように設定する方法も考えられる。3) 更に本手法は，音響モデルを用いずに発声中に含まれる音素相当のまとまりを捉え，その数量に基づいて評価を行なっている．そのため，英語のシャドーイングを行なう際に日本語を発声しても，未知語を適当に発声しても，高い評価が得られることになる．今後，HMM に基づく自動評定技術とのハイブリッド化も含め，このような問題にも対処する必要がある．4) 現在，言語教育の立場からシャドーイングを眺めた際，韻律的特徴に基づくシャドーイングも注目されてきている．本自動評定手法においてもピッチ，イントネーションといった韻律情報を用いた評価を積極的に取り入れていく必要性があるだろう．

文　献

[1] 門田修平, “シャドーイングと音読の科学”, コスモピア株式会社, 2007.

[2] 玉井健, “リスニング指導法としてのシャドーイングの効果に関する研究”, 神戸大学大学院総合人間科学研究科博士学位論文, 2001.

[3] 下村直也他, “制約条件つきクラスタリングによる連続音声からのイベント境界検出”, 信学技報, SP2007-12, pp.25–30, 2007.

[4] S. Dusan *et al.*, “On the relation between maximum spectral transition positions and phone boundaries,” Proc. InterSpeech, pp.17–21, 2006.

[5] Y. P. Estevan *et al.*, “Finding maximum margin segments in speech,” Proc. ICASSP, pp.937–940, 2007.

[6] Y. Qiao *et al.*, “Unsupervised optimal phoneme segmentation: objectives, algorithm and comparisons,” Proc. ICASSP, 2008 (to appear).

[7] 宮本定明, “クラスター分析入門 ファジィクラスタリングの理論と応用”, 森北出版, 1999.

[8] N. Minematsu *et al.*, “Para-linguistic information represented as distortion of the acoustic universal structure in speech,” Proc. ICASSP, vol.1, pp.261–264, 2006.

[9] M. Nakamura *et al.*, “Acoustic and linguistic characterization of spontaneous speech,” Proc. Int. Workshop on Speech Recognition and Intrinsic Variations, pp.3–8, 2006.