

英語リスニング中の脳血流量を用いた正則付回帰モデルによる 言語能力推定法

亀山 拓朗 鏑木 崇史 栗原 陽介
 青山学院大学 理工学部

1 はじめに

英語は現在，世界共通語となりつつある．非英語圏である日本においても，文部科学省がスーパーグローバル大学事業を施策[1]するなど，国際的に活躍できる人材の育成に力を入れており，第二言語として英語習得の重要性は強く認識されている[2]．英語習得における学習では，リーディング，ライティング，リスニング，スピーキングの4項目に大別されている．特にリスニングやスピーキングはリアルタイムでのコミュニケーションの必要性から重要視されている．英語習熟度の評価において近年特に英語でコミュニケーションをとるといふ観点からの評価が重要視されてきているが，TOEICなどの試験は時間がかかる点や結果が来るまでに時間がかかることなどの問題が生じる．脳活動を使った日本人の第二言語習得に関する研究として，篠塚による第二言語習得の考察などがあるが，これらの研究は英語コミュニケーションの中でもスピーキングに着目している[3]．リスニングに着目した研究として，筆者らはこれまで英語習熟度レベルを評価した研究を行ってきた[4]．

本論では前者が着目したスピーキングではなくコミュニケーションにおいて必要とされているリスニングに着目し，ワーキングメモリの箇所，すなわち前頭葉の箇所の脳血流量から英語のリスニングにおける能力を評価する方法を提案する．

2 提案手法

本研究ではNIRSにより測定した脳血流量を用いて英語リスニングにおける言語能力を推定するシステムを提案する．図1に本研究で提案するシステムの概要を示す．

リスニング中において測定した酸素化ヘモグロビン(Oxy-Hb)，脱酸素化ヘモグロビン(Deoxy-Hb)，総ヘモグロビン(Oxy-HbとDeoxy-Hbの和)の変化量を $x_n(k)$ と置く．ただし n はNIRSチャンネル番号， k は離散時間である．それぞれの平均，

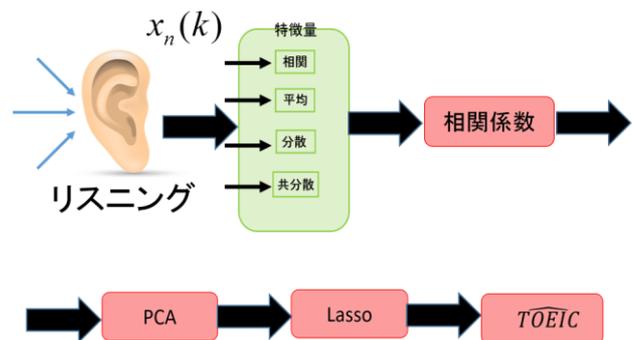


図1 実験の概要

分散，共分散，相関を求めると，

$$\bar{x}_n = \frac{x_n(1) + x_n(2) + \dots + x_n(k)}{k}$$

$$\sigma_n^2 = \frac{\sum \{x_n(k) - \bar{x}_n(k)\}^2}{k}$$

$$S_{mm} = \sum \{x_n(k) - \bar{x}_n(k)\} \{x_m(k) - \bar{x}_m(k)\}$$

$$r_{nm} = \frac{\sum \{x_n(k) - \bar{x}_n(k)\} \{x_m(k) - \bar{x}_m(k)\}}{\sqrt{\sum \{x_n(k) - \bar{x}_n(k)\}^2} \sqrt{\sum \{x_m(k) - \bar{x}_m(k)\}^2}}$$

となる．特徴量 \mathbf{a} は上記の基本統計量を並べたベクトルとなる．

$$\mathbf{a} = (\bar{x}_n, \sigma_n^2, S_{mm}, r_{nm})$$

リファレンスとしての英語習熟度スコア E を用いて特徴量 \mathbf{a} の各成分との相関を求める．求めた相関係数のうち，絶対値が閾値 T を上回るもののみを寄与特徴量 \mathbf{b} として選択する．寄与特徴量 \mathbf{b} に対し主成分分析を行い，主成分特徴量 \mathbf{c} を計算する．

英語習熟度スコア E と主成分特徴量 \mathbf{c} を用いて正則付回帰を行い，回帰係数 $\hat{\beta}$ を求める．

英語習熟度が未知の脳血流量 $x_n^{new}(k)$ は，同様の処理で求めた主成分特徴量 \mathbf{c}^{new} と回帰係数 $\hat{\beta}$ を用いて

$$\hat{E} = \hat{\beta} \mathbf{c}^{new} + \beta_0$$

として求める．

3 実験

CH7~CH16 まで測定できる日立メディコ製の光トポグラフィを使用し、眼窩前頭皮質を計測対象とする。被験者は21-23歳の男女11人に協力してもらった。実験自体は被験者に目を閉じてイヤホンを入れてもらい、英語のニュースを3トピックス流す。課題時間は1分間。課題の前後で安静状態を測れるように、課題の前で約30秒、後で約30秒の安静時間を設ける。安静状態は体に制限をかけない状態のことを指している。また、事前知識がないために脳血流量が反応しないことを未然に防ぐために、スキーマを与えてから英語のニュースを使いリスニングの実験を行った。リスニングをきちんと行ってもらうためリスニング後に簡単なテストを設け、理解度の確認を行った。テストは被験者に事前に伝えてある。リファレンスとしてはTOEIC®を使用し、提案手法の妥当性を検討する。

評価は11名のデータを交差検定で行い、推定したTOEICスコア \hat{E} と実際のスコア E^{true} の絶対誤差の平均を求める。さらに、寄与特徴量 b を求めるときの閾値 T の影響と主成分分析の有無における有効性について検証する。

4 実験結果

図2に閾値 T と平均誤差との関係を示す。閾値 $T=0$ のとき、 $a=b$ となり、この実験結果から相関を用いてある程度相関のある特徴量のみを抽出することが有効であることが分かる。さらに、主成分分析の有無により平均誤差が大きく異なることも分かった。

一番平均誤差が小さいときの値は89.93となった。表1に実際のTOEICスコアと推定スコアを示す。TOEICスコアが低い被験者と高い被験者においては誤差が大きくなる傾向が観測できた。

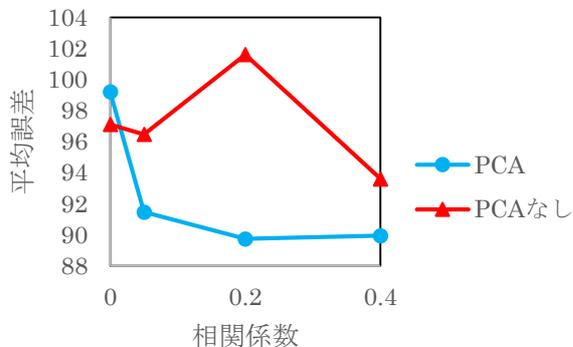


図2 閾値 T と平均誤差との関係

表1 推定スコアと誤差

	TOEIC スコア	推定スコア	誤差
ID 1	420	582	-162
ID 2	480	630	-150
ID 3	520	541	-21
ID 4	590	620	-30
ID 5	600	604	-4
ID 6	660	651	9
ID 7	680	627	53
ID 8	685	629	56
ID 9	710	612	98
ID 10	720	665	55
ID 11	730	667	63

5 考察

主成分分析を行い、関係性の高い特徴量をまとめたことにより、正則付線形回帰を行った際に関係性の低い特徴量をノイズ成分として推定式から排除できたことが考えられる。

一方、TOEICスコアの低い被験者と高い被験者の誤差が大きいことが目立った。これは今回の被験者はTOEICスコアの分布に制限を設けずに取得したため、データ数の多いTOEICスコア中央付近で精度が高くなることが理由と考えられる。

今後はTOEICスコアの分布が一様になるような被験者母集団を用いて推定することで、さらに精度が高まることが期待できる。

参考文献

- [1] 文部科学省：“スーパーグローバル大学事業”(2014)
- [2] 文部科学省：“「英語が使える日本人」の育成のための行動計画”(2003)
- [3] 篠塚勝正：“言語脳科学に基づく第二言語習得の考察”成城英文学第32号(2008)
- [4] 安中悠介ほか：“眼窩前頭皮質における脳血流量による言語習熟度指標の提案”電子情報通信学会(2015)

A Language Skill Estimation Method from Cerebral Blood Flow while English Listening Comprehension Using Regularized Regression Model

†Takuro Kameyama †Takashi Kaburagi †Yosuke Kurihara Aoyama Gakuin University