

研究種目：基盤研究（C）
 研究期間：2007～2009
 課題番号：19500799
 研究課題名（和文） 合成音声を利用した外国語リスニング能力測定支援システムの開発と実践的評価
 研究課題名（英文） Development and practical evaluation of support system for measuring listening comprehension and utilizing synthetic speech in foreign language learning
 研究代表者
 康 敏 (KANG MIN)
 神戸大学・大学院国際文化学研究科・准教授
 研究者番号：60290425

研究成果の概要（和文）：本研究では、外国語の教育現場において手軽に学生のリスニング能力を測定して学習効果を確認する学習支援システムを開発した。リスニング能力測定教材を容易に作成するために、合成音声を利用できるようにした。自然音声との差を考慮して、個々の単語の合成音声使用の可否に関して学生による合成音声評価結果を元に、評価スコアを提示することにした。また測定データに対する分析手法を提案して実践的評価によってシステムの利用がリスニングに対する弱点の発見につながる事が判明した。

研究成果の概要（英文）：In this research, we have developed and practically assessed a web-based system for foreign language learning to support the measurement of listening comprehension. A text-to-speech engine is combined so that teachers can easily create their own web-based listening tests. Because there still is a gap between natural speech and synthetic speech in naturalness and sometimes in intelligibility, the system provides evaluation scores of voices according to the evaluation result of synthetic speech by students to help teachers determine if a voice can be used in listening test or not. Moreover, we invented an approach to analyze the measurement data and clarified in the practical assessment of the system that the approach is effective for identifying weak points in listening.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2007年度	1,600,000	480,000	2,080,000
2008年度	600,000	180,000	780,000
2009年度	600,000	180,000	780,000
年度			
年度			
総計	2,800,000	840,000	3,640,000

研究分野：総合領域

科研費の分科・細目：科学教育・教育工学・教育工学

キーワード：e-ラーニング 音声合成技術 外国語教育 システム開発 リスニング能力

1. 研究開始当初の背景

外国語によるコミュニケーション能力が

強く求められる現在、教育現場においてはリスニングとスピーキングを中心とした、より

実践的なカリキュラムが展開されてきている。効率的な語学学習には、音声や映像教材が欠かせないため、音声機器や映像機器を使用する LL 授業が一般的に行われてきた。しかし、各機器に対応した教材をそれぞれ用意する煩わしさだけでなく、教師が各機器の操作方法を把握していないと使用できず、また把握していてもそれぞれの機器を使う準備がその都度必要なため、授業時間のロスが大きい、という問題も出てきたのである。この問題点を克服するものとして登場したのが CALL である。インターネットの普及につれ、近年、ほとんどの CALL システムはクライアント側のブラウザ経由でサーバ側にある教材を利用するネットワーク型になりつつある。いつでもどこでも利用できるのがこの種の CALL システムの大きな特徴である。従来の授業と比べ、開発した教材はより共有しやすくなるだけでなく、学習者にバラエティある授業形態を提供することも可能である。

しかし、多くの CALL システムにおいては、いったん開発が終わると、教材の内容が固定されがちであり、教師側による手軽な変更ができない。そのため、市販のオーサリングツールを利用して教師自身による教材開発が多く見られるようになったが、LL 授業と同様に、音声の録音と編集、画像の録画と編集などの作業が必要であるため、授業の準備時間が長くなる問題が依然として残っている。さらに、システムへの理解やコンピュータに関する一定の専門知識なども求められる。これはまた CALL 授業が LL 授業のように展開できない要因の一つとなっている。最近、手軽にオンライン教材を作成できる市販の web オーサリングツールも出されているが、クライアント側にもソフトが必要であったり、語学学習専用でないため、CALL プラットフォームで利用する際、テンプレートが柔軟性に欠けていたりするなどの問題が指摘されている。コストをかけて導入した CALL 教室の利用率は期待ほど高くないのが現状である。

このように、CALL 授業を展開していくには、これらの問題を解決した教材開発システムが望まれる。その解決策の1つとして合成音声の利用が考えられる。音声合成技術の急速な進歩に伴って、現段階の合成音声は明瞭さ (Intelligibility) だけでなく、自然さ (Naturalness) においてもかなり自然音声に近づいてきている。市販の電子辞書の多くはすでに音声合成機能が実装されている。しかし、合成音声の教育への利用の有効性と実用化についての研究は非常に少ないため、合成音声を利用した外国語学習システムに関する実践的報告がほとんどない。筆者らはこの点に着目し、これまでに英語と中国語を対象にして合成音声の学習者・教師評価を調査し、

合成音声の外国語教育への利用について研究を行ってきた。その成果の一部として、単語の音韻的構成にもよるが、現段階の合成音声の一部は学習者にとって自然音声 (ネイティブ音声) と差がないことを明らかにした。

2. 研究の目的

本研究の目的は、合成音声を利用した外国語リスニング能力測定支援システムを開発し、その実践的評価と利用によってシステムの有効性を確かめると共に教育現場において手軽に学生のリスニング能力を測定して学習効果を確認できる CALL システムの実用化を目指すことである。また開発したシステムに対する実践的評価を通じて、学習者のリスニング能力測定結果を蓄積・分析し、学習者の音声に対する認知的特徴を探ること、データマイニング手法などを用いてシステムによる学習者の外国語リスニング能力測定データ分析を行い、その結果を学習者と教師にフィードバックできる知的支援システムの実現も目標としている。

音声合成技術をシステムに取り入れることによって、教師が手軽にリスニング能力測定する教材を作成できるのみならず、学習者のリスニング能力測定結果に基づいて学習者のニーズにあったリスニング学習教材の自動生成も可能となり、ネイティブスピーカーが少ない環境での CALL 教材作成に多くの可能性を持たせることができる。

3. 研究の方法

(1) システムはサーバ・クライアント型システムとし、リスニング能力測定用教材作成を支援するシステムと学習者のリスニング能力を測定してその結果を蓄積・分析するシステムに分かれる。開発プログラム言語として主に PHP を使用し、開発環境として、Web サーバには Apache、DBMS には、MySQL を使用して、TTS エンジンには AT&T Natural Voices を利用する。

(2) 教材作成支援システムでは、教師が Web 型インターフェースを通じて測定教材を作成する。測定教材作成にはリスニング学習に特化したシステム用意のテンプレートが用いられ、テキスト文章を入力できるようにする。測定教材の音声については音声合成エンジンによる作成と既存の音声ファイルアップロードという2種類の方法が選択できる。学習者のリスニング能力を測定してその結果を蓄積・分析するシステムでは、テスト問題のプロトタイプは、音声を聞いてその書き取りと和訳を回答するというスタイルである。学習者が測定教材の問題に答える以外に、聞いた音声の音質 (聞き取りやすさ) について「はっきり聞き取れる」、「聞き取れる」、「ど

ちらとも言えない」、「聞き取りにくい」及び「まったく聞き取れない」という5段階評価をしてもらうことにより、システムの利用に伴って合成音声の学習者評価を蓄積・分析し、合成音声の評価結果データベースの改良を図る。

(3) 音声合成エンジンによって合成音声を生成し、教材に取り入れる場合、システムはこれまでの合成音声評価研究の結果を元に、評価スコアを算出して、合成音声を利用可能かどうかのための参考基準として教師に提示する。

(4) 測定及び結果の蓄積・分析システムでは従来の採点方式で回答者の正答率を計算する以外に、データマイニング手法を取り入れ、学習者のリスニング特性の推測を行い、その結果をフィードバックする。

4. 研究成果

(1) 平成19年度では、単語レベルのリスニング能力測定システムの開発を行った。教材作成支援システムでは、教師はWeb型インターフェースを通じて測定教材を作成する。インターフェースには、多肢位選択形式のテンプレートを用意した。これまでの合成音声評価研究では、音韻的特徴によって、自然音声と同じ、または自然音声より聞き取りやすい合成音声が決して少なくないことを明らかにしている。本研究では、個々の合成音声の利用可否について判断するため、測定教材作成に合成音声を取り入れる際、これまでの評価結果を元に、その合成音声に関する評価スコアを提示するアプローチを開発した。その詳細は以下に示す。

評価スコアの計算は測定教材に使う音声より下位レベルで行われる。単語の場合は音素や音節などとなる。単語に対する音声知覚は、単語の下位の音声知覚単位に依存すると判明されているため、単語の合成音声評価はその下位の知覚単位の影響を反映するものであり、評価スコアはこれらの単位に対する評価から算出すべきと考えられるが、現段階においてその基本単位がまだ不明瞭である。そのため、本システムでは、音素を利用することとした。単語の発音を音素に分解するにはCMU発音辞書を使った。また、直接に音素に対する評価を行うのが困難であり、その評価は単語またはセンテンスに対する評価データから推測することにした。その推測にベイズ統計を用いた。

推測に使用する合成音声評価結果データベースは評価テーブル、集計テーブル及び音素テーブルから構成される。表1に示しているように、五段階評価レベルは評価データとしてテーブルの「評価レベル」フィールドに

格納され、それぞれの単語に対応している。集計テーブルには評価テーブルから単語の各評価レベルの数をカウントして格納している。音素テーブルには発音辞書と評価テーブルに基づいて、すでに評価されたことのある音素の一覧をリストアップしている。評価テーブルの「単語」は音素テーブルの「単語」とリレーション関係にある。新しい評価結果が評価テーブルに追加される場合、音素テーブルも更新される。

表1 合成音声評価結果データベースの例

ID	評価者	属性	評価レベル	単語
1	Student1	1	3	look
2	Student2	1	4	look
3	Student3	1	4	look
4	:	:	:	:
	:	:	:	:

*評価テーブル

単語	評価レベル	集計	音素	単語
look	1	1	L	look
look	2	8	UH	look
look	3	17	K	look
look	4	6	K	break
:	:	:	:	:

*集計クエリ

*音素テーブル

単語の評価スコアはその単語を構成する各音素の各評価レベルにおける期待値の平均と定義し、音素の期待値は、ベイズ統計に従って推定を行う。ある音素がある評価レベルに評価される確率は、その評価レベルにおける事後確率になり、ベイズ定理によって求められる。

$$P(i / Ph_j) = \frac{P(Ph_j / i)P(i)}{\sum_{m=1}^n P(Ph_j / m)P(m)}$$

ここに、 i は i 番目の評価レベルを表し、 $1, 2, 3, \dots, n$ の中から一つの値を取りうる。五段階評価の場合、 $n=5$ である。 Ph_j は j 番目の音素を意味している。 $P(m)$ は m 番目の評価レベルの確率で、集計テーブルから評価レベル m のデータ総数対データ総数の比によって決められる。 $P(Ph_j / m)$ は音素 Ph_j の評価レベル m における事前確率で、音素テーブルから得られたその音素が対応するすべての単語の m におけるデータ総数対評価レベル m のデータ総数の比によって求められる。そして、各音素の期待値を計算して単語の評価スコアはその平均値となる。音素テーブルにない音素については、各評価レベルにおける事前確率 $P(i / Ph_j)$ ($i=1, 2, \dots, n$) が

同じと仮定して、五段階評価の場合は 1/5 となる。

(2) 平成 20 年度では、昨年度に開発した単語レベルのリスニング能力測定システムに、ベイズ統計による評価スコアの提示機能をシステムに実装して提示アルゴリズムの有効性について分析を行った。提示アルゴリズムを検証するにあたって、“receive”と“theory”を例にして計算を行った。“receive”に関しては計算したスコアと実際の評価値の間に差が小さいが、“theory”に関しては、大きな差が生じている。この違いをもたらす理由はいくつか考えられる。まず単語において音声知覚の基本単位が音素とする仮定にある。“theory”においての差はこの仮定に問題があると示唆するものと考えられる。次に、音素が基本単位と前提する場合、各音素の単語の音声知覚に及ぼす影響が異なることも考えられる。つまり、単語の評価スコアを算出するとき、各音素の期待値の単純平均ではなく、各音素の単語の音声知覚への影響を考慮した加重平均を用いる必要がある。また、検証に使用したデータが少なく、評価データによらない音素があったことも、結果の精度に影響した原因と考えられる。

(3) 平成 21 年度では、これまでの計 2844 件の評価データをシステムのデータベースに格納し、合成音声評価スコアの提示アルゴリズムによる計算結果を推測値にするとともに、新しい合成音声を作成してシステムの実践的評価を行った。評価スコア提示アルゴリズムの有効性を確認し、測定データに対する分析手法を提案して実践的評価によってシステムの利用がリスニングに対する弱点の発見につながる事が判明した。

これまでの評価データからより長い音節の単語または短母音 /ɪ/, /ʊ/ と /ə/, 長母音 /i:/, /u:/ と /ɜ: r/ 及び二重母音 /aɪ/, /ɔɪ/ を含む単語の合成音声は比較的高く評価される傾向が見られ、摩擦音においては、種類によって評価が分かれることが判明している。この結果を参考にして、提案アルゴリズムの実践的評価では、長母音 /i:/, /u:/ を含む単語や摩擦音がある単語を選んだ。被験者は 37 名の大学一年生である。サンプル音声は長母音を含む単語群と摩擦音中心の単語群に分かれる。被験者にはこれらの音声を聞き、そのスペルと和訳を回答欄に記入して、聞きやすさについて五段階評価 (1-5) を行ってもらった。これらの実践的評価データを測定値として評価スコア提示アルゴリズムによる推測値と比較した。長母音を含む単語群のうち、5 割の単語については、測定値と推測値の差が ±0.2 以内であり、8 割強に

ついてはその差が ±0.9 以内に収まっている。差の絶対値が 1 を超えたものはわずか 2 割弱である。摩擦音中心の単語群については、ほぼ同様な結果が得られている。従って、本研究で提案した評価スコア提示アルゴリズムは有効であると考えられる。一方、推測値と測定値の差が大きくなる理由について、更なる分析が必要である。理由の一つとして、一部の音素には評価データがなく、一律に 1/5 の確率を使用していることが挙げられる。また単語の親密度によって評価にもたらすバラツキも要素として取り入れる必要がある。

測定データからリスニングに対する弱点、特に特定の音韻に対する知覚の弱点を見つけ出すには、評価スコア提示アプローチと同様に、ベイズ統計を用いた。分析対象となる測定データは、研究方法で述べてプロトタイプ問題の場合、書き取りが正解か不正解のデータとなる。分析は一人の学習者においても一学習グループにおいても同様に行うことができる。単語に対して低い正解率が出たとき、これは単語を構成するある音素またはいくつかの音素に対する知覚に困難が見られるからという考えはキーポイントである。ここで、正解した単語を音素に分解して各々の音素を集合 A に、不正解の単語を分解して各々の音素を集合 B に属しているとする。集合 A と集合 B の要素はそれぞれ音素テーブルにある正解と不正解の音素に対応する。測定データによって同じ音素であっても、同時に集合 A と集合 B に属する場合がある。集合 A は音素の聞き取りに問題がない (正解) という事象 A に対応し、集合 B は音素の聞き取りに問題がある (不正解) という事象 B に対応している。ある音素に対して聞き取りに問題があるについての推定はその音素を含む単語を聞いたとき不正解となる事後確率の算出によって行われる。この事後確率が大きい場合は、その音素の聞き取りに困難があると推測できる。

ベイズの定理によって、ある音素 Ph において聞き取りに問題がある事後確率の算出は以下となる。

$$P(B/Ph) = \frac{P(Ph/B)P(B)}{\sum_{m=A,B} P(Ph/m)P(m)}$$

ここでは、P(A) と P(B) は事象 A と事象 B の確率であり、これまで学習したすべての単語において正解した単語の割合と不正解の単語の割合を計算して得られる。一学習者の場合は、その学習者に関するこれまでのすべての測定データが処理対象となるが、一学習グループの場合は、そのグループに属するすべてのデータが対象となる。P(Ph/A) と P(Ph/B) は条件付き確率であり、それぞれ正

解と不正解の単語の中に音素 Ph を含む単語の割合と定義される。

上述の手法を用いて計42名の大学一年生に単語聞き取り訓練テストを行い、得たデータを分析した。その結果、母音/i:/の不正解率が高いことが分かった。テストに用いた単語の中、母音/i:/の不正解率に寄与したのは”keep”, ”leave” と”speak”であるが、単語”keep” と”speak”の正解率が高くそれぞれ0.86と1.00である。母音/i:/の高い不正解率に大きく寄与したのは単語”leave”である。”leave”は単語としての正解率はわずか0.02である。間違った回答のほとんどでは”leave”を”live”としている。母音/i:/に対する知覚、または母音/i:/と/l/の区別ができていないことが推測できる。また、学生の”leave”に対する「聞き取りやすさ」に関する5段階評価の平均値は2.9であり、聞き取りにくいことが分かる。このように、ベイズ統計による測定データに対する分析手法をシステムに取り入れることよって測定データから音声知覚に関する問題点を発見できることが明らかである。特にデータが大量の場合、システムの利用によってリスニングに対する弱点の発見がしやすくなる。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計4件)

- ① 康敏、柏木治美、鏑木誠、合成音声評価スコアを提示するリスニング教材作成システムの開発について、日本教育工学研究報告集、査読無、JSET09-1、2009、pp. 365-368
- ② Kanetsugu Sakata、Min Kang、Kazuo Asakimori、Makoto Kaburagi、An Analysis for Student Opinion Survey with Item Response Theory、Proc. of 8th International Conference on Information Technology Based Higher Education and Training、査読有、2009、403-407
- ③ 柏木治美、康敏、大月一弘、外国語学習における音声合成利用の現状と課題、神戸大学国際コミュニケーションセンター論集、査読無、No. 5、2008、pp. 11-20
- ④ Min Kang、Harumi Kashiwagi、Jutta Treviranus、Makoto Kaburagi、Synthetic speech in foreign language learning: an evaluation by learners、International Journal of Speech Technology、査読有、Vol. 11、No. 2、2008、pp97-106

[学会発表] (計2件)

- ① 康敏、単語認知に影響する音声要素分析のための学習支援システム、日本教育工学会第25回全国大会、2009年9月20日、東京大学
- ② 柏木治美、合成音による英単語の聞き取りに関する基礎的調査、日本教育工学会第25回全国大会、2009年9月20日、東京大学

6. 研究組織

(1) 研究代表者

康 敏 (KANG MIN)

神戸大学・大学院国際文化科学研究科・准教授
研究者番号：60290425

(2) 研究分担者

柏木 治美 (KASHIWAGI HARUMI)

神戸大学・国際コミュニケーションセンター・
准教授
研究者番号：60343349

鏑木 誠 (KABURAGI MAKOTO)

神戸大学・大学院国際文化科学研究科・名誉教授
研究者番号：40093504