様式 C-19

科学研究費補助金研究成果報告書

平成２３年５月３１日現在

機関番号：１７４０１
研究種目：基盤研究（C）
研究期間：2008～2010
課題番号：２０５２０４７２
研究課題名（和文）　学習者の音環境に影響を受け難い第二言語習得システム

研究課題名（英文）　2nd language learning system without dependence on sound environment

研究代表者
茁木　禎史（CHISAKI YOSHIFUMI）
熊本大学・大学院自然科学研究科・准教授
研究者番号：50284740

研究成果の概要（和文）：
CALL 教室など複数の発話学習者が同時発話をした際，隣接話者の発話の影響により正確な自動評価ができない。これは学習意欲を低下させるため，現状では正しい発話評価を行えるようにヘッドセットを用いている。本課題では，ヘッドセットを用いずに学習意欲を持続する発話訓練システムを実現することを目的とする。これを実現するために，各学習端末で得られる隣接話者の発話信号をネットワーク経由で送信し，NTP の時刻に基づく同期処理を施したそれらの信号を用いて対象話者の発話評価精度を向上させる。

研究成果の概要（英文）：
Efficiency on utterance learning of 2nd language is degraded when multiple learners exist in CALL room due to the simultaneous utterances by other learners. A purpose of the research is to improve the performance on utterance evaluation when the simultaneous utterances exist for sustainable learning. The system enhances a target learner’s utterance using signals of other learner’s utterances and multi channel synchronization technique based on network time protocol.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2008 年度	2，100，000	630，000	2，730，000
2009 年度	900，000	270，000	1，170，000
2010 年度	500，000	150，000	650，000
総　計	3，500，000	1，050，000	4，550，000

研究分野：　音響情報学
科研費の分科・細目：言語学・日本語教育
キーワード：言語学習システム，e-learning，雑音抑圧，非同期信号，多点観測，マイクロフォンアレー

１．研究開始当初の背景

言語学習において，文法のみならず，聞き取りと発話の訓練が必須であり，多くの言語教育者がマンツーマンの発話訓練に十分な時間を割くべきであると認識している。しかし，学習者の数に対して教師の数が十分でなく，授業時間内に学習者個々人への対応が難しい。発話訓練は授業中であれば限られた学習者の発話を教師が評価して正すことができるが，自学自習の際には発話の評価者がおらず，発話の習得が困難である。授業中にマンツーマンで行われることが理想的である発話訓練に対する不満は，教師，学習者の両者の抱える大きな問題である。

我々は昨今の日本への留学生を指導する留学生センター教員らと連携し，日本語を第

二言語として習得するための初級レベルのコンテンツを有する発話訓練システムの開発を行っている。

現場での発話訓練に対する不満の解決方法の一つとして，様々な研究者が自動発話訓練機能を有する e-learning システムの開発を試みており，急激な進歩がみられる。一般に，e-learning を利用する学習者に対して，適切で正確なポジティブフィードバックを与えることの有効性は，既に下記の文科省「特色ある大学教育支援プログラム」による熊本大学全学での取り組みによって示されている。発話訓練機能のある e-learning システムにおいても同様に適切なフィードバックを与えることは，学習者の意欲喪失を防ぐことや学習効率の向上を実現するために必須である。自動発話訓練システムでは，適切なフィードバックを学習者へ与えるためにシステムが正確に学習者の発話を理解する必要がある。一般に，発話評価には自動音声認識システムを活用するが，現状では学習者が満足できる品質ではない。その結果，学習意欲が向上しないことは既に研究報告されている。

音声認識の性能低下には，大きく二つの原因が考えられ，一つは，音声認識システムで学習している辞書や音響モデルが適切でないために生じる認識率の低下である。もう一つは，雑音による認識率の低下である。前者の問題は，例えば，韓国語を母語とする学習者は，「つ」が母語にないために韓国語においてそれに最も近い「ちゅ」と発話している。具体的には，「ひとつ」を「ひとちゅ」と発話し，音声認識システムが第二言語学習者特有の発話に適応できていないことである。我々は既に音声認識データベースに学習者が間違えやすい単語セットを加え，誤った発話の適切な評価が行えるように，さらには自動音声認識の音響モデルも発話訓練用に改善し，認識率の向上を実現している。一方，後者の雑音による影響を低減させるためにヘッドセットを用いて認識率を維持しているのが現状である。

しかし，学習者がヘッドセットを身につけなくて済むハンズフリー音声認識を要求していることは，ハンズフリー機能が既にテレビ会議システムやコンシューマープロダクトである携帯電話に実装され，利用されていることから明らかである。

学習端末でハンズフリー機能を実現するには，自宅学習の際は学習端末から再生される教示音声と環境音や室内の反射の影響が，音声認識率低下を招き，学習者の意欲が低下する。さらには，教室のように学習端末が多数隣接する状況では，隣接する学習者の発話が本来の学習者の音声と混ざるため，認識率が低下する。これらの音環境の影響を低減することを考慮した e-learning システムが十分に検討されていないことは普及を妨げる一因であり，解決すべき大きな課題である。

２．研究の目的

本課題では，自学自習が効率的に行える発話の評価機能を有する言語習得システムの高性能化を目的とする。特に，発話の自動評価に必須である自動音声認識の性能を，先端音声処理技術と次世代ネットワーク技術の融合により改善することである。

学習者の音環境に左右されない効率的な発話訓練システムを実現するには，音環境の能動制御，あるいは集音時における音響信号処理が考えられる。音環境を制御する方法の一つに，能動的にスピーカーから逆位相の音を放射する方法があるが一般的な生活環境においては複数のスピーカーが必要であり，多くの学習端末が必要となる e-learning システムでは現実的ではない。また，ロボットとの対話などを目的とした自動音声認識システムでは，複数話者の同時発話による，もしくは環境雑音による音声認識率の低下は，複数のマイクロフォンを等間隔に並べたマイクロフォンアレーにより避けられるが，十分な性能を得るには多数が必要であり，現実的ではない。

隣接する学習端末を活用することは，多点観測による情報量増加により対象音源信号の強調には大きく貢献する。しかし，それぞれの学習端末で音響情報の時刻が同期されている必要がある。その同期を実現するために，ネットワーク上で音響情報を共有する際には，パケット毎に時刻も含む情報が必要である。また，音響信号を扱うにあたって，アナログ信号をデジタル信号に変換し，さらにはそのデータをネットワークカードからパケット送出し，受け取るまでの遅延も問題である。しかし，昨今のハードウェアにおいても絶対的な AD 変換の遅延や認識処理による遅延があるものの，高性能化により学習者が耐えることができる程度の遅延が期待できるため，同期が取れれば発話訓練に活用できる。よって，本課題では，分散配置された学習端末からの非同期多チャンネル音響信号の同期をとり，隣接端末から得られた音響信号から対象の音声信号を強調する技術の実現を目指す。この技術が実現されれば，学習者が隣接する，もしくは定常状態に近い環境雑音がある状況での音声認識率向上につながり，e-learning システムでの言語習得効率の向上が期待できる。

３．研究の方法

本研究の目的であるネットワークを経由した多チャンネル非同期信号を用いた音声強調手法の開発を行うに際し，①利用環境における隣接話者の発話混入の度合い，②

チャンネル間の時刻同期方法，③既存の信号強調手法に対する時刻同期誤差の影響および認識性能の 3 点について検討を行った。

① 利用環境における隣接話者の発話の影響

残響のある CALL 教室での配置において，認識対象話者に対して前方正面，前方左右，話者の左右に関して音圧計測を行った。

② チャンネル間の時刻同期方法

チャンネル間の時刻同期には，信号から時刻差を推定する方法と外部時刻同期技術を用いた方法を検討する。前者は各端末で取得した音声信号のパケットだけで済むが，後者は各端末で端末内の時刻を同期することと送信パケットにタイムコードを埋め込む必要がある。前者の観測信号から時刻差を推定する手法では，信号の相関およびコヒーレンスに基づき処理をするが，実時間処理を行うためのフレーム長では十分な性能が期待できない。一方，後者の手法は，ネットワークタイムプロトコル(NTP)による時刻同期処理の負荷があるが前者に比べ演算量が少なく誤差が小さいことが期待できる。これら 2 つの手法について検討を行う。

③ 既存の信号強調手法に対する時刻同期誤差の影響

音声強調手法には，信号の統計的な情報により分離する優れた手法があるが，十分な性能を得るために事前にある程度の信号長が必要となる。それは学習者に発話評価結果を示すまでの応答時間に直接影響し，システムの利用持続を低下させるほどの処理時間を繰り返すことになる。また，事前に雑音推定を行って推定した雑音を差し引くスペクトラルサブトラクション法も考えられる。この手法は，事前に雑音の推定が必要であるため，CALL 教室での利用や隣接話者の発話が動的に変化する状況では十分な性能を発揮できない。そこで，処理時間が短いバイナリマスク法を用いて，ある程度の雑音低減が可能であることをシミュレーションで確認する。

４．研究成果

①に関しては，一般的な CALL 教室において対象話者と隣接話者が最も近い状況として 0.8m の間隔を想定した。この状況において，対象話者と隣接話者が同時発話をした際に，隣接話者（妨害音）のパワーが対象話者の音声信号のパワーより大きくなることはほとんど無いことが確認でき，シミュレーションにおいては希に生じる悪い条件として 0dB を設定することとした。

②に関しては，各端末で得られる非同期信号の相互相関およびコヒーレンスを用いた手法は，演算量が多くかつ同期精度が十分でないことが確認できた。外部時刻同期を用いる方法では， NTP を各端末で動作させ，各端末での取得信号を集約して強調処理を行うサーバーにおける同期処理時間が±2ms 以下の誤差を生じることが確認できた。

③に関しては，②の NTP を用いた手法の時刻誤差を含む同期処理済みマルチチャンネル信号を用いてシミュレーションを行った。その雑音低減性能は音声認識による評価を行った。その結果の一部を以下に示す。

図 1 は，残響のある学生居室で最大 6 人の学習者が同時に SNR=0dB で発話した際の発話評価結果である。灰色が同期処理および強調処理を適用する前，黒色が本研究で開発した手法を適用した単語認識結果である。

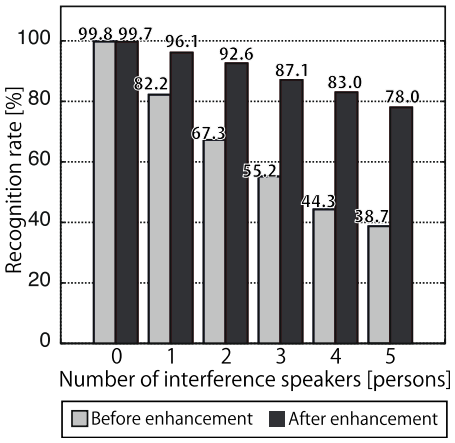

図 1 同時発話人数に関する認識率

この結果では，すべての学習者がヘッドセットを利用せずに，対象話者以外の学習者 2 名が同時に発話した際は処理前が 55.2%であるが，開発した手法を用いると 87.1%まで改善しており，有効であることが示された。この結果は，同時発話かつその SNR が 0dB という最も悪い条件であり，一般的な利用においては，学習意欲を削ぐまでの誤認識が生じないことが期待できる結果となった。

図 2 は，残響のある学生居室でいくつかの妨害話者の配置を組み合わせた際の発話評価結果である。SNR は 0dB である。結果の横軸における I は右隣もしくは左隣のいずれか一人が対象話者と同時に発話した場合，II は対象話者に向かって前方の学習者が同時に発話した場合である。また，III は右斜め向かいの学習者どちらかが発話した場合，IV は，両隣の学習者 2 名が同時に発話した場合，V は，右斜め向かいの学習者 2 名が同時に発話した場合である。実環境において，両隣の学習者が対象話者と同時に同じパワーで発話しても，ヘッドセットを用いなくても 52.5%から 85.5%まで改善している。この結果は，両隣の学習者が同時に同じパワーレベルで発話することが希であることを鑑みると，学習意欲を削ぐまでの誤認識が生じないことが期待できる結果となった。

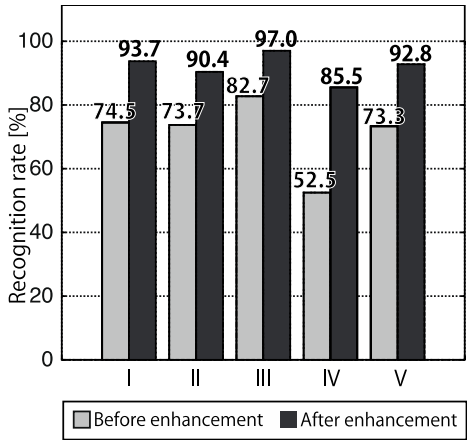

図 2　妨害話者の位置と認識率

これらの結果から，認識率の低下は，対象話者と他の学習者（妨害話者）の距離に依存していることが明らかになった。

このように，開発した手法の有効性が示せたため，実時間処理に関しても検討を行った。学生が入手可能な性能を有するラップトップ PC で実装を行い，学習者が発話してからその評価を示すまでの処理時間を計測した。結果，その処理時間は 550ms を必要とすることがわかり，学習意欲を削ぐような応答速度ではないことが示された。

本課題では，発話評価の性能が学習意欲を低下させることに着目し，発話評価の精度向上を目指した。学習端末の発達により，NTP の利用が現実的である状況において，十分に実時間処理が可能な手法を実現した。低価格な PC においても十分な応答速度を実現し，学習意欲を削ぐことないレベルでの発話評価精度が得られた。

今後は，バイナリマスク法という単純な雑音低減手法の代わりに効果的なマスク方法を導入することによりさらなる認識性能の改善を検討する必要がある。また，大規模な学習教室においてもネットワーク負荷の影響による応答時間の遅延が学習効率にどのような影響を与えるかの検討を要する。

５．主な発表論文等

〔学会発表〕（計７件）

1) 苣木禎史，眞島智久，宇佐川毅，Network time protocol 同期を用いた非同期多チャンネル信号の音声強調 -複数妨害音が音声認識性能へ与える影響 -，日本音響学会 2011 年春季研究発表会講演論文集，pp.1-4(CDROM)，2011.3.9～11，東京，早稲田大学

2) Tomohisa Mashima, Yoshifumi Chisaki, Tsuyoshi Usagawa, Speech Enhancement Method utilizing Asynchronous Signals with Time Code over TCP/IP Network, Proc. Asia Pacific Signal and Information Processing Association, p.13(CD-ROM), 2010.12.14 ～ 17, Singapore, Biopolis

3) Tomohisa Mashima, Yoshifumi Chisaki, Tsuyoshi Usagawa, Speech enhancement method based on spectral filtering utilizing asynchronous signals with embedded timecode over TCP/IP based network, Proc. TENCON2010, pp.1341-1346, 2010.11.21 ～ 24, Fukuoka, International Convention Center

4) 眞島 智久，苣木 禎史，宇佐川 毅，非同期マルチチャネル信号を用いた単語音声認識のための信号強調 - NTP によるチャネル間同期法の検討 -，電子情報通信学会技術研究報告 110(285)，pp.73-78，2010.11.19，福岡，九州大学

5) 眞島 智久，金 佳英，苣木 禎史，宇佐川 毅，TCP/IP ネットワークを介した非同期マルチチャネル信号を用いた雑音低減手法の検討 - 雑音推定誤差と自動音声認識への影響 -，IEICE 電子情報通信学会 信号処理研究専門委員会 第 24 回信号処理シンポジウム，講演論文集 CD-ROM，pp.198-202，2009.11.25 ～ 27，鹿児島，鹿児島サンロイヤルホテル

6) Tomohisa Mashima, Kousuke Matsuo, Yoshifumi Chisaki, Tsuyoshi Usagawa, Spectral subtraction method utilizing asynchronous signals observed over TCP/IP based network, Proc. The 10th Western Pacific Acoustics Conference, pp.1-6 (CD-ROM), 2009.9.21 ～ 23, Beijing, China, Beijing Friendship Hotel

7) Kousuke Matsuo, Yoshifumi Chisaki, Tsuyoshi Usagawa, A spectral subtraction method using signals observed at separate computers connected by network, Proc. Youngnam-Kyushu Joint Conference on Acoustics 2009, pp.135-138, 2009.2.7, Andong, Korea, House of Korean Culture

６．研究組織

(1)研究代表者
苣木　禎史（CHISAKI YOSHIFUMI）
熊本大学・大学院自然科学研究科・准教授
研究者番号：50284740

(2)研究分担者
宇佐川 毅（USAGAWA TSUYOSHI）
熊本大学・大学院自然科学研究科・教授
研究者番号：30160229