

機関番号：33934

研究種目：基盤研究 (C)

研究期間：2008～2010

課題番号：20500177

研究課題名 (和文) 装着型および環境設置型マイクロホンを用いた実環境下での音声認識

研究課題名 (英文) Speech Recognition Using Wearable and Environmental Microphones under Real Environments

研究代表者

實廣 貴敏 (JITSUHIRO TAKATOSHI)

愛知工科大学・工学部・准教授

研究者番号：60394996

研究成果の概要 (和文)：2 種類の異なるマイクロホンを用い、自動で時刻同期しながら雑音観測用のマイク入力を認識対象用のマイク入力から引き去ることで、雑音抑圧を行う方法を提案した。2 つのマイク入力の時刻同期を音声特徴量の距離を比較することで行い、また、自動推定される変換行列により、空間特性を補正しつつ、雑音抑圧できる。実験結果から、特に変動する雑音に効果があることがわかった。

研究成果の概要 (英文)：We propose a new noise suppression using two kinds of microphones with auto time-synchronization. The one of microphones is used to estimate noise signals, and the estimated noise components are subtracted from the target input signal of the other microphone. Auto time-synchronization is done by using speech features' distance, and the space characteristics of noise spectrum is compensated by an estimated conversion matrix. Experimental results show that this method is more effective for non-stationary noise signals.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008 年度	1,300,000	390,000	1,690,000
2009 年度	400,000	120,000	520,000
2010 年度	600,000	180,000	780,000
総計	2,300,000	690,000	2,990,000

研究分野：音声情報処理

科研費の分科・細目：情報学・知覚情報処理・知能ロボティクス

キーワード：音声認識, 雑音抑圧, ウェアラブルコンピューティング

1. 研究開始当初の背景

病院での看護師らの作業など、複数名での作業の管理を、ウェアラブルコンピューティングとセンサーネットワーク技術を用いて行うことを考える。その一つとして、音声認識技術を用いる。ユーザの胸元などに装着したマイクロホン (装着マイク) を通し、体に装着した小型端末で、作業内容や現状の状況などを音声で録音し、無線 LAN 経由でデータを管理するサーバへ送信する。その音声を音声認識技術により認識、システムで各ユーザの作業状況を把握する。これにより、予定の

作業進捗を管理したり、問題が生じていることを検出したりすることができる。

しかし、このような音声認識では、胸元に付けたマイクから周辺の雑音が混入する。また、ユーザは動き回るため、場所や作業の状況により、観測される雑音が異なる。これらの雑音を抑圧することは、容易ではなかった。

2. 研究の目的

1. で説明したような状況下での音声認識における、変動する雑音を抑圧する手法として、雑音観測用のマイクロホン入力を併用し、

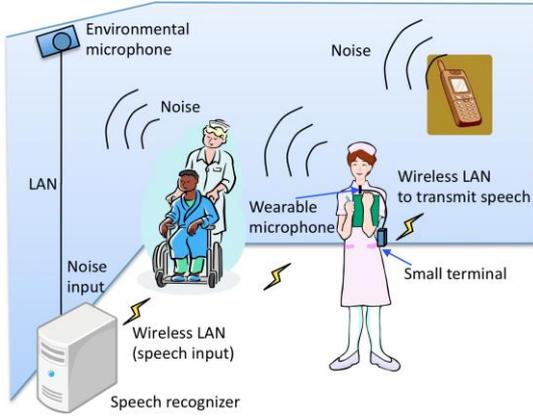


図 1 想定している状況の例

自動で時刻同期を取りながら、雑音を減算する手法を提案した。想定している例を図 11 に示す。1. のような状況下で、ユーザが活動する場所、各部屋などに雑音観測用マイクロホン（環境マイク）を設置し、装着マイクとは別に雑音を観測するものとする。環境マイクの入力もサーバへネットワーク経由で送信する。サーバでは、装着マイクおよび環境マイクの 2 つの時刻同期が取れていない入力を自動で同期しつつ、雑音抑圧を行い、音声認識に役立てる。なお、ユーザの位置検出が同時に必要だが、赤外線通過センサを用いるなど、別の手段を利用することにする。

3. 研究の方法

まず、2 つのマイク入力を用いて雑音抑圧が可能かどうかを検証する。次に、時刻同期が取れていない 2 つの入力を自動で同期し、雑音抑圧が可能か検証する。

(1) 環境音の周波数特性に対する補正

2 種類のマイク入力での空間特性およびマイク特性から同じ雑音であっても観測される周波数特性が変化する。これを補正する。

装着マイクの第 i フレーム入力スペクトル・ベクトル X_i 、含まれる音声 S_i 、雑音 N_i とすると書きのように表される。

$$X_i = S_i + N_i$$

ω を周波数として、一般的に Spectral Subtraction (SS) 法による推定音声は、推定された雑音 $\hat{N}_i(\omega)$ を用いて、

$$\hat{S}_i(\omega) = \{|X_i(\omega)| - |\hat{N}_i(\omega)|\} e^{j \arg X_i(\omega)}$$

と表される。 $\arg X_i(\omega)$ は位相成分である。よく使われる SS 法では、定数 α を掛け、調整する。

$$\hat{N}_i = \alpha \bar{N}$$

\bar{N} は発話直前の非音声区間の平均がよく使われる。これは定常雑音を仮定していることになり、SS 法は非定常雑音には弱い。

そこで、本提案法では、環境マイクの入力 Y_i を雑音成分と見なし、時刻同期された各フレームの Y_i を用いることで非定常雑音にも対処する。 \bar{N} を Y_i に置き換えた提案法 SS2mic は下記のようになる。

$$\hat{N}_i = \alpha Y_i$$

さらに、変換行列をかける場合を提案する。

$$\hat{N}_i = A Y_i$$

変換行列 A は対角行列を用い、各対角成分は N フレームの 2 つの入力の比とする。

$$a(\omega) = \frac{\sum_{i=1}^N |X_i(\omega)|}{\sum_{i=1}^N |Y_i(\omega)|}$$

これにより、空間特性の違いも吸収できる。推定に用いる N フレームは SS と同様、発話直前の区間や別に観測された雑音区間から抽出する。

(2) 装着マイクと環境マイク間の環境音および音声に対する同期方法

2 つのマイク入力の音声特徴量空間での比較を行い、最も近くなる時刻で同期をとる。今回は、2 入力の音声特徴量間のユークリッド距離を用いた。装着マイク入力の第 n フレームの p 次元特徴量ベクトルを $x_n = (x_{n1}, x_{n2}, \dots, x_{np})^T$ 、環境マイク入力の第 m フレームの特徴ベクトルを $y_m = (y_{m1}, y_{m2}, \dots, y_{mp})^T$ とする (T は転置)。フレームのずれを τ としたときのユークリッド距離を下記のように定義する。

$$D(\tau) = \frac{1}{N - \tau} \sum_{n=1}^{N-\tau} \sum_{m=\tau+1}^N d(n, m)$$

$$d(n, m) = \sqrt{\sum_{i=1}^p (x_{ni} - y_{mi})^2}$$

ここで、 N は使用する総フレーム数、 $D(\tau)$ を最小とする τ を同期する位置とする。

4. 研究成果

(1) 実験条件

評価データおよび雑音データともあらかじめ用意しておき、図 2 に示す収録環境で再収録することで、装着および環境マイク双方で得られるデータを作成する。今回、話者は移動なしで、環境マイクから 1 m あるいは 2 m 離れているとした。

アルゴリズム自体はタスクに依存しないため、タスクに依存しない評価データおよび

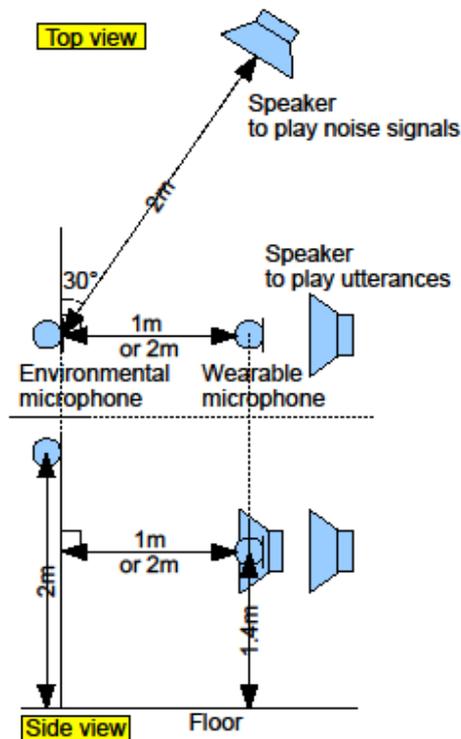


図 2 評価データ収録条件

雑音データを用いた。評価用音声データは日本音響学会新聞記事読み上げ音声コーパスから IPA-98-TestSet を用いた。男女各 23 名による計 200 発話であった。雑音データとしては、校舎建築工事、および学園祭で収録されたデータを用いた。工事雑音は定常雑音に重機の突発的な雑音がたびたび挿入されたもの、学園祭雑音は音楽が始終流れている中で頻りに様々な人声が挿入された変動雑音である。装着マイクに audio-technica 製 AT9642、環境マイクに DPA 製 4060 を用いた。収録機器として装着型に Roland 製 R-09HR、環境設置型に同じく R-44 を用いた。48 kHz サンプルング 16 bit で収録後、16 kHz サンプルングにダウンサンプルングした。フレーム長 25 ms、フレーム周期 10 ms を用いた。雑音抑圧は 512 点 FFT を行った上で、波形へ逆変換した後、音声認識を行った。従来法 SS では発話直前の 100 ms を雑音成分とした。提案法 (SS2mic) の変換行列は同じ区間から推定した。音声認識の特徴量は MFCC, Δ MFCC, Δ 対数パワーおよび CMS を用いた。語彙数は約 2 万語、音響モデルは不特定話者 2000 状態 16 混合分布 HMM, 言語モデルは 75 ヶ月分の新聞記事で学習された bigram および trigram であった。音声認識エンジンとしては Julius ver. 4.1 を用いた。

(3) 時刻同期が既知のときの提案雑音抑圧手法の評価

まず、時刻同期ができていないとして、雑音抑圧の基本的な検証を行った。作成した評価

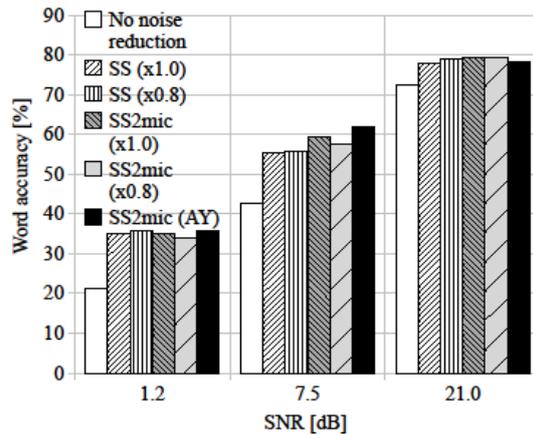


図 3 工事雑音に対する評価

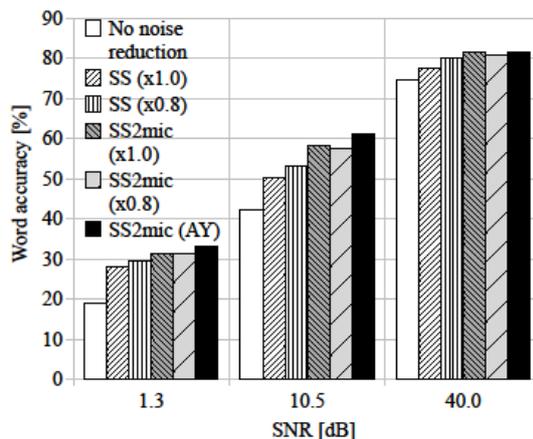


図 4 学園祭雑音に対する評価

データ中、マイク間の距離が 1 m のときの実験結果として工事雑音および学園祭雑音に対する単語認識率を図 3, 4 に示す。図中、“(x1.0)”, “(x0.8)” とあるのは SS の係数で、比較的良好な性能であった 0.8 を掲載した。提案法 “SS2mic (AY)” は変換行列を用いた場合である。提案法は全般的に工事雑音に比べ、学園祭雑音では改善度が高い。非定常雑音に対し、特に効果があることが分かる。

(4) 時刻同期を行いながら、雑音抑圧を行う提案雑音抑圧手法の評価

工事雑音に対する 1 m 離れている場合の単語認識率を図 5 に、同様に、学園祭雑音に対する場合の単語認識率を図 6 に示す。図中、“SS” は従来法、“SS2mic” は手動同期の提案法、“SS2mic (auto)” は自動同期の提案法である。同様に、2 m 離れている場合の実験も行った。得られた傾向は同じであったので、ここでは省略する。

詳細な結果は省略するが、自動同期による同期時刻の推定誤差は一部を除き、 $\pm 10 \sim 20$ ms 程度であった。時折、 $\pm 100 \sim 300$ ms 程度の誤差が得られた場合でも、もともと高 SNR であることが多く、雑音抑圧後、音声認識の結果としては、大きな影響は見られなかった。

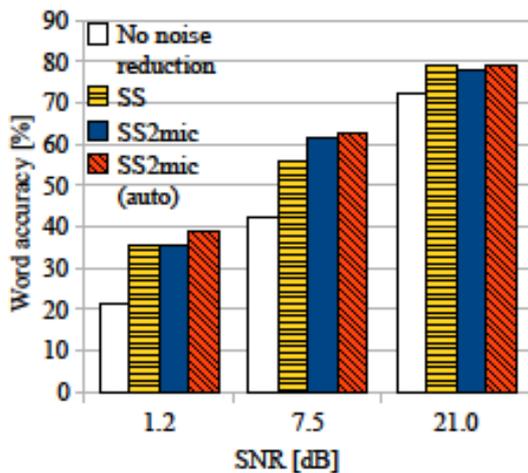


図 5 工事雑音に対する結果 (SS2mic: 時刻同期なし, SS2mic (auto): 時刻同期あり)

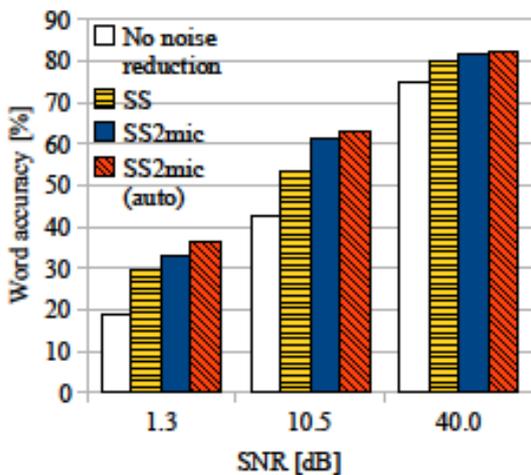


図 6 学園祭雑音に対する結果 (SS2mic: 時刻同期なし, SS2mic (auto): 時刻同期あり)

自動同期の音声認識精度は手動同期と同じか、やや良い結果になった。誤り改善率は工事雑音の一部を除き、10~20%が得られた。全自動で雑音抑圧でき、従来法より良い性能が得られることが分かった。

(5) 研究成果のまとめ

本研究では、2種類の異なるマイクロホンを用い、自動で時刻同期しながら雑音観測用のマイク入力を認識対象用のマイク入力から引き去ることで、雑音抑圧を行う方法を提案した。環境マイクにより雑音を推定し、その入力と認識対象の装着マイク入力との時刻同期を音声特徴量の距離を比較することで行った。雑音成分を引き去る際には、空間特性を補正するために、周波数スペクトルの特性を自動で推定される変換行列を用いた。実験結果により、変動する雑音に対しても有効に働く雑音抑圧手法とすることが可能であることが分かった。今後の課題としては、環境マイクをより活用し、入力から観測され

る頻度の高い音源を自動でモデリングし、さらにそれを雑音抑圧などに利用する、などが考えられる。

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0件)
特になし

[学会発表] (計2件)

① 真野雄大, 五十右準, 實廣貴敏, 小暮潔, “2種類のマイクロホンを利用して自動同期しつつ雑音抑圧する手法,” 日本音響学会 2011年春季研究発表会講演論文集, pp.155-156, 2011年, 査読なし

② 實廣貴敏, 小暮潔, “装着型および環境設置型マイクロホンを用いた音声認識のための雑音抑圧手法,” 日本音響学会 2009年春季研究発表会講演論文集, pp.151-152, 2009年, 査読なし

[図書] (計0件)
特になし

[産業財産権]

○出願状況 (計0件)
特になし

○取得状況 (計0件)
特になし

[その他]

○ホームページ

愛知工科大学工学部情報メディア学科
實廣研究室ホームページ
<http://www1.aut.ac.jp/~jtlab/>
「研究資金」のページ
本研究の概要あり

6. 研究組織

(1) 研究代表者

實廣 貴敏 (JITSUHIRO TAKATOSHI)
愛知工科大学・工学部・准教授
研究者番号: 60394996

(2) 研究分担者

なし

(3) 連携研究者

小暮 潔 (KOGURE KIYOSHI)
金沢工業大学・工学部・教授
研究者番号: 50395159