様式 C-19

科学研究費補助金研究成果報告書

平成　22 年　5 月　1 日現在

研究種目：　若手研究(B)
研究期間：　2007～2009
課題番号：　19700163
研究課題名（和文）実環境バイモーダル音声認識共通評価基盤の構築
研究課題名(英文) Audio-visual speech corpus for evaluating speech recognition performance in noisy environments

研究代表者
宮島　千代美（MIYAJIMA CHIYOMI）
名古屋大学・大学院情報科学研究科・助教
研究者番号：90335092

研究成果の概要（和文）：車内雑音環境での音声認識を対象としたバイモーダル音声認識評価用のコーパスを構築するため，室内と車内でバイモーダル音声データを収録しデータ整備を行った．室内データについては，車内で収録した雑音と映像の輝度変化を利用して雑音を重畳することにより，車内の雑音環境を模擬した．また，評価実験を通じて，評価の基準となるベースラインの特徴量や認識条件を選定し，評価用スクリプトやマニュアルと併せて DVD にて配布する準備が整った．これらは，研究用データベースとして，今後配布される予定である．

研究成果の概要（英文）：Audio-visual speech data are collected in a silent room and a vehicle for developing an audio-visual speech corpus which is used for evaluating speech recognition performance in noisy environments, especially in in-car environments. Acoustic noise and gamma values of images are used for simulating in-car environments over the recorded data in the silent room. Baseline audio and visual features and an integration method are calibrated in some experimental evaluations. The corpus will be open to the public along with database manuals for research purposes.

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	0	1,600,000
2008 年度	1,100,000	330,000	1,430,000
2009 年度	625,880	187,764	813,644
年度			
年度			
総　計	3,325,880	517,764	3,843,644

研究分野：総合領域
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：バイモーダル音声認識，データベース，雑音環境，車内雑音

１．研究開始当初の背景

音声入力インタフェースは，ボタンやタッチパネルに触れることなく，ハンズフリー・アイズフリーの入力操作が可能であり，カーナビのような車載システムのインタフェースとして，安全運転支援の観点において利用価値が高いと言える．しかし，実走行車内では，ロードノイズやエンジン音，音楽や同乗者の話し声といった背景雑音の影響で，雑音を音声として誤検出する「発話区間検出誤り」や，発話内容を誤認識する「音声認識誤り」が起こるため，まだ利用者が満足できる音声インタフェースを提供するには十分に至っていない．

走行車内のような雑音環境に対してロバストな音声認識の１つのアプローチとして，雑音の影響を受けない映像情報を音声情報に併用して認識を行うバイモーダル音声認識の研究が広く行われている．しかし，実際にドライバが運転をしながら音声インタフェースを利用するような状況での認識実験は広く行われておらず，車内音声認識を目的としたバイモーダル音声認識の研究用データベースで公開されているものは見当たらない．

２．研究の目的

本研究では，雑音環境下での音声認識，特に，車内で音声認識インタフェースを利用する状況下での音声認識性能を評価するために，走行車内でドライバが発話する音声と顔映像を収録し，バイモーダル音声認識の研究用コーパスを構築する．また，室内で収録したバイモーダル音声データを音声認識用モデルの学習に利用することを想定し，室内環境で収録した音声と映像に事後的に雑音を重畳した，車内雑音環境シミュレーションデータを作成し，学習・評価用コーパスとして公開する．これにより，研究者が同じ土俵でバイモーダル音声認識技術の性能を比較評価ができると考えられる．また，構築したデータベースを利用して，比較の対象となるベースラインの特徴抽出・統合方法について検討する．

３．研究の方法

(1) バイモーダル音声データの収録

音声と映像は，図１のような同期収録システムを構成して収録する．音声は２つのマイク，顔映像はカラーカメラと近赤外カメラで収録し，メディアコンバータでステレオ音声とカラー映像，およびステレオ音声と近赤外映像を統合し，それぞれDVフォーマットのAVIファイルとして保存する．

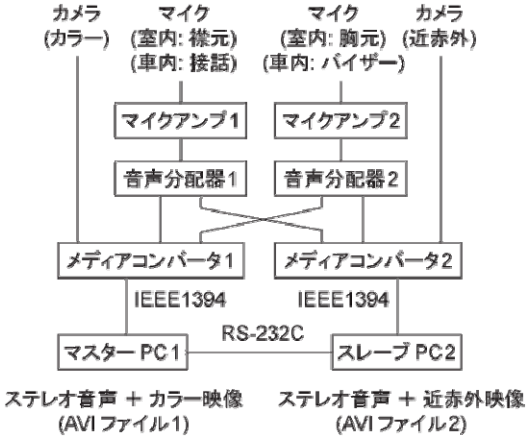


図 1: 収録システムの概要

車内の雑音環境としては，データ収録車において，ダッシュボードとステアリングコラムにカラーカメラと近赤外カメラをそれぞれ取り付け，走行環境として，アイドリング・市街地走行・高速走行の3種類，車内環境として，通常・エアコンON・音楽再生・窓開けの4種類の組み合わせで計12種類の条件で収録を行う．室内および自動車内で収録した音声・映像を元に，バイモーダル音声認識評価研究用のコーパスを構築する．

(2) 雑音重畳によるシミュレーション

CENSRECシリーズの雑下音音声認識評価用コーパスの構築基準に則り，市街地および高速道路走行時の乗用車雑音を，室内のクリーンな環境で収録した音声にさまざまなSN 比で重畳することで，車内雑音環境を模擬したデータベースを構築する．なお，重畳する車内雑音は，ドライバの襟元に音声収録に用いたものと同じマイクを装着し，ドライバが発話をし

ない状況で収録を行う.

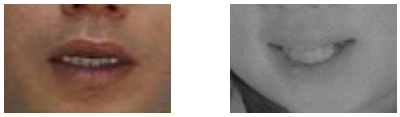
映像については，走行中の乗用車内で収録した映像を元に，室内で収録した口唇映像の輝度をガンマ補正することにより，車内雑音環境を模擬する.

(3) データベースの構造

コーパスには，音声およびカラーカメラと近赤外線カメラの二種類の映像から，口唇付近の画像をのみ切り出したデータを収めるとともに，隠れマルコフモデルツールキットをベースとした認識評価用のスクリプトやデータベースのマニュアル等も一緒に収め，認識性能の比較評価が容易にできるようにする.

４．研究成果

まず，バイモーダル音声データベースの整備を進め,図2の例に示すような口唇動画像と音声を含む学習・評価データセットを作成した.



(a)カラー画像 (b) 近赤外画像
図2：口唇画像の例

音声認識評価実験を通じて，ベースラインの特徴量や統合方法について検討し，音声特徴量は，メルフィルタバンクケプストラム係数とその動的特徴量，画像特徴量は，口唇画像の主成分得点とその動的特徴量をベースラインの特徴量とすることとした．また，音声と映像の特徴量をマルチストリームの隠れマルコフモデルで初期統合に基づいてモデル化し，ストリーム重みを変化させ，最も認識性能が高くなるストリーム重みで統合した場合をベースラインとすることとした．データベースの利用者は，このベースラインの認識率と比較することで，利用者が提案するバイモーダル音声認識の特徴抽出手法や認識手法の性能を評価できる仕組みとなっている.また,本コーパスは，バイモーダル音声データと併せて，雑音の重畳や，音声・画像の特徴抽出のプログラム，および無償提供されている隠れマルコフモデルツールキットＨＴＫに準じた音声認識評価用のスクリプト，およびデータベース利用のためのマニュアルも同時に収められていることから，バイモーダル音声認識や発話区間検出の研究において広く利用されると期待される.本コーパスは，今後研究用として公開・配布される予定である.

５．主な発表論文等

〔雑誌論文〕（計３件）

① N. Kitaoka, T. Yamada, S. Tsuge, C. Miyajima, K. Yamamoto, T. Nishiura, M. Nakayama, Y. Denda, M. Fujimoto, T. Takiguchi, S. Tamura, S. Matsuda, T. Ogawa, S. Kuroiwa, K. Takeda, and S. Nakamura, CENSREC-1-C: An evaluation framework for voice activity detection under noisy environments, Acoustical Science and Technology, 査読有, pp.363-371, 2009 年.

② 二宮芳樹，坂義秀，前野俊希，根木大輔，宮島千代美，森健策，北坂孝幸，末永康仁，音声と画像の統合によるドライバの発話区間検出，映像情報メディア学会誌，査読有，vol.62, no.3, pp.435-441, 2008 年.

③ 原直，宮島千代美，伊藤克亘，武田一哉，多様な音響環境下における音声認識システム利用時のデータ収集システム，電子情報通信学会論文誌，査読有，vol.J90-D, no.10, pp.1115-1123, 2007 年.

〔学会発表〕（計１０件）

① 田村哲嗣，宮島千代美，北岡教英，武田一哉，山田武志，滝口哲也，柘植覚，山本一公，西浦敬信，中山雅人，傳田遊亀，藤本雅清，松田繁樹，小川哲司，黒岩眞吾,中村哲，CENS-REC-1-AV: マルチモーダル音声認識コーパスの構築，2010 年日本音響学会春季研究発表会，2010 年 3 月，調布市.

② 伊藤新，原直，宮島千代美，北岡教英，武田一哉，複数音響モデルからの最適選択による音声認識，2009 年電気関係学会東海支部連合大会，2009 年

9月，豊田市.

③ 武田一哉，尾崎晃，マルタルーカス，西脇由博，宮島千代美，北岡教英，自動車運転コーパスにおける行動観測信号の統合と利用，2009年マルチメディア，分散，協調とモバイルシンポジウム，2009年7月，別府市.

④ S. Tamura, C. Miyajima, N. Kitaoka, S. Hayamizu, K. Takeda, CENSREC-AV: Evaluation frameworks for audio-visual speech recognition, 2008 International Conference on Auditory and Visual Speech Processing, 2008年9月，オーストラリア.

⑤ M. Nakayama, T. Nishiura, Y. Denda, N. Kitaoka, K. Yamamoto, T. Yamada, S. Tsuge, C. Miyajima, M. Fujimoto, T. Takiguchi, S. Tamura, T. Ogawa, S. Matsuda, S. Kuroiwa, K. Takeda, S. Nakamura, CENSREC-4: Development of evaluation framework for distant-talking speech recognition under reverberant environments, 2008 International Conference on Spoken Language Processing, 2008年9月，オーストラリア.

⑥ L. Malta, P. Angkititrakul, C. Miyajima, K. Takeda, In-car speech data collection along with various multimodal signals, 2008 IEEE Intelligent Vehicles Symposium, 2008年2008年6月，オランダ.

⑦ A. Ozaki, S. Hara, T. Kusakawa, C. Miyajima, T. Nishino, N. Kitaoka, K. Itou, K. Takeda, In-car speech data collection along with various multimodal signals, 2008 Language Resources and Evaluation Conference, 2008年5月，モロッコ.

⑧ T. Nishiura, M. Nakayama, Y. Denda, N. Kitaoka, K. Yamamoto, T. Yamada, S. Tsuge, C. Miyajima, M. Fujimoto, T. Takiguchi, S. Tamura, S. Kuroiwa, K. Takeda, and S. Nakamura, CENSREC-4: Development of evaluation framework for distant-talking speech recognition under reverberant environments, 2008 Language Resources and Evaluation Conference, 2008年5月，モロッコ.

⑨ N. Kitaoka, K. Yamamoto, T. Kusamizu, S. Nakagawa, T. Yamada, S. Tsuge, C. Miyajima, T. Nishiura, M. Nakayama, Y. Denda, M. Fujimoto, T. Takiguchi, S. Tamura, S. Kuroiwa, K. Takeda, and S. Nakamura, Development of VAD evaluation framework CENSREC-1-C and investigation of relationship between VAD and speech recognition performance, 2007 IEEE workshop on Automatic Speech Recognition and Understanding, 2007年12月，京都市.

⑩ C. Miyajima, T. Kusakawa, T. Nishino, N. Kitaoka, K. Itou, and K. Takeda, On-going data collection for driving behavior signal, 2007 Biennial on DSP for in-Vehicle and Mobile Systems, 2007年6月，トルコ.

〔図書〕（計１件）

① S. Tamura and C. Miyajima, Multimodal Speech Corpora for Robust Japanese Speech Recognition in Noisy Environments, S. Itahashi and C.Y. Tseng eds., Computer Processing of Asian Spoken Languages, Section 4.9 (3), 5 pages, 2010.

６．研究組織

(1)研究代表者

宮島　千代美（MIYAJIMA CHIYOMI）
名古屋大学・大学院情報科学研究科・助教
研究者番号：90335092