「多点観測音声処理の研究」

名古屋大学大学院 情報科学研究科 メディア科学専攻　教授

武田 一哉

1　はじめに

統計的な音声認識手法の高い基本性能にも関わらず，音声認識はいまだ多くの応用分野で利用されるに至ってはいない。その原因の全てを技術的な問題に帰することはできないが，最も重要な問題の一つとして，「実際の利用時における認識性能を制御することが困難である」点が挙げられる。本稿では，多点音響観測に基づく車内音声認識の高度化について報告する。

2　音声収録機器と収録内容

実際の走行環境下における音声を収集するために，実験車を作成した。車内には計測用のPCが搭載されるとともに，分散して（運転者と実験ナビゲータの接話マイク，左右ダッシュボード，左右バイザー，天井（２箇所），マイクロホンアレー（４マイク，運転席バイザー位置））マイクロホンが配置されている。収録内容は (1) 音素バランス文（７５文／人），(2) 車内音声対話（４０分／人），(3) 孤立単語，連続数字（１００発声／人），の３種類であり，８００名の運転者についてデータが収集されている。車内音声対話は，同一のタスク（レストラン案内）について，人間との対話，WOZ システムとの対話，音声対話システムとの対話，の３種類を収集し，発声内容にタグを付与している。

ここでは本コーパスを用いた研究例として，分散して設置されたマイクを用いて対数スペクトル領域で重回帰を行うことで，接話マイクのスペクトルを近似する雑音抑圧法を紹介する。

この方法は，

$$\log|X_0(k)| = \sum_{i=1}^{N} w_i(k) \log|X_i(k)|, \qquad (1)$$

に従って，分散マイクスペクトル X_i と回帰重み w_i により，接話マイクのスペクトル X_0 を近似し，認識に利用する。この手法では，w_i を適切に設定できれば遠隔マイクの誤り率を半減させることができるが，最適な回帰重みは走行条件（停車／市街地走行／高速走行，窓開／オーディオ／エアコン等）に依存する。そこで分散マイクで受音された信号のスペクトルに基づいて，走行時の車内の音環境を予めいくつかのクラスに分類し，クラスごとに最適な回帰重みを学習した。認識時には，無音区間の分散マイクのスペクトルから車内の音環境のクラスを推定し，当該クラスに最適な回帰重みを選択すればよい。

この方法で得られた性能と，接話マイク受音，遠隔マイクによる受音，とを比較した結果を図１に示す。自動選択された，回帰重みを用いた場合でも，走行条件が与えられた場合と同等の性能が得られることが分かる。この結果は，いくつかの環境毎に最適化されたシステムを選択的に利用することで，様々な環境下で動作可能なシステムが構築可能なことを示唆している。

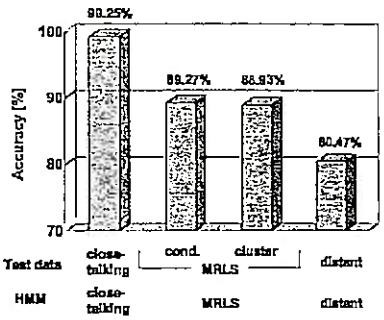


図 1: 分散マイクによる対数スペクトル重回帰法の性能（孤立５０単語の認識）。MRLS-cond: 走行環境を既知として走行環境毎に最適化された重回帰重みを用いた場合，MRLS-cluster: 分散マイクの受音パワーから走行環境を推定して，重回帰重みを適応的に用いた場合。

参考文献

[1] 河口信夫，岩博之，牛窪誠一，武田一哉，稲垣康善，板倉文忠，“車内音声対話収集システムの開発” 電子情報通信学会論文誌 DII, Vol.J84-DII, No.6, (2001.6) pp.909-916

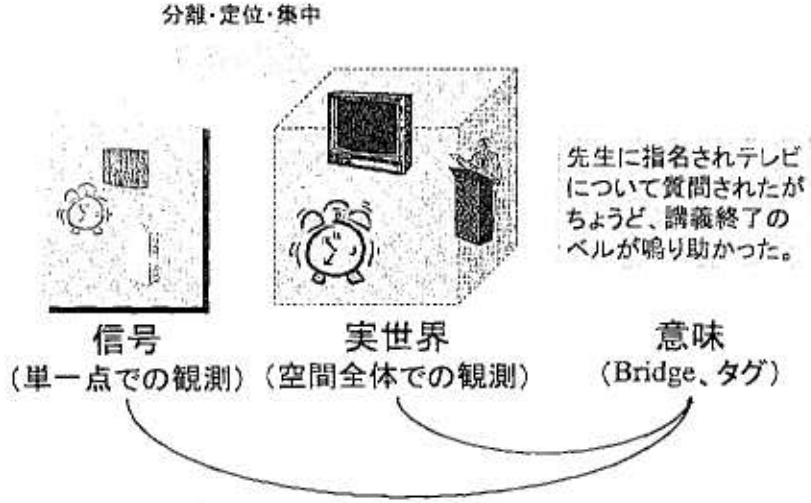
*Speech processing for distributed multiple signal capturing by Kazuya Takeda, (Nagoya University, takeda@is.nagoya-u.ac.jp)

多点観測音声処理の研究

武田一哉

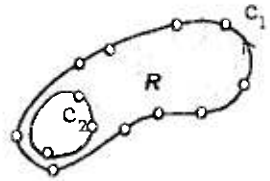
- 多点(多重)観測と知的統合
  - 空間補間、情報統合
- 分散マイクロホン音声処理
  - ダイバーシティ、統合、適応
- 空間物理音響
  - 空間音響再生
- 実世界データベースと競争的評価
  - A proposal for International Alliance for Advanced Studies on In-Car Human Behavioral Signals

実世界の多点観測と知的統合

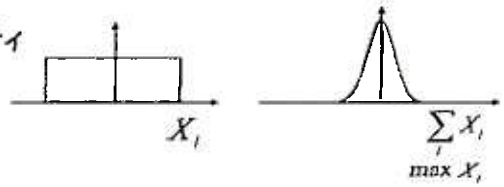


多点(多重)観測

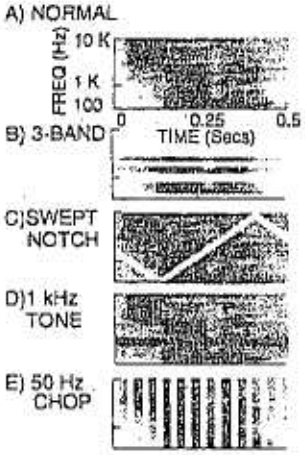
- 空間:多点観測
  - サンプリング、補間
  - 音場再現



- 情報:多重観測
  - ROVER (Recognizer Output Voting Error Reduction)
  - Missing feature
  - 空間ダイバーシティ



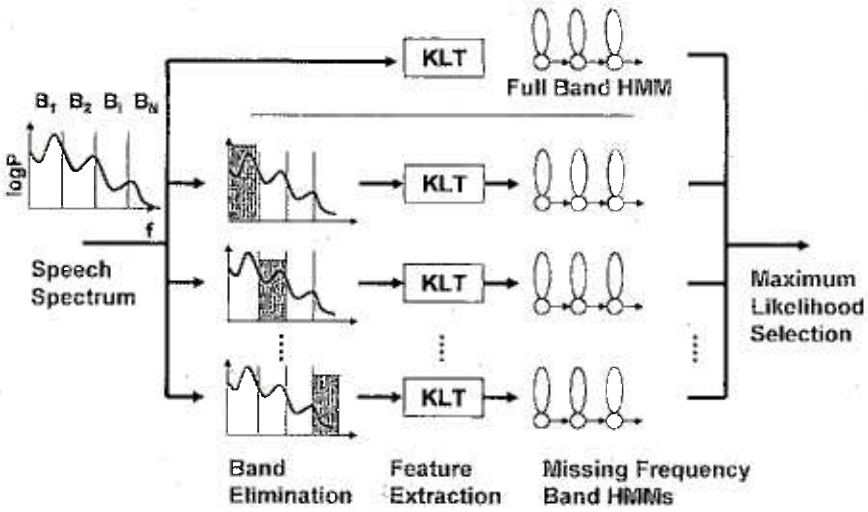
Missing feature theory



REF. MFB	REF. CEP.	MISS MFB
98.8%	99.2%	98.8%
9.1%	14.3%	93.3%
14.2%	9.7%	98.9%
55.9%	11.6%	97.0%
9.1%	82.6%	96.6%

Lippmann and Carlson, ISCA Workshop on ASRU 1997

Missing featureの工学的実現



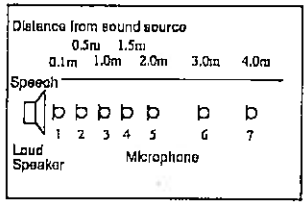
対話音声からの意味抽出

- 車内音声対話の例
  - （正解）：あー 手持ち の お金 が ない から ちょっと 銀行 の キャッシュコーナー 寄り たい けど 東海銀行 の キャッシュコーナー は 近く に ある か な 。
  - （認識結果）：そこ 手持ち の お金 が ない から ちょっと 銀行 の チャーシュー ほう が いい ん だ けど 東海銀行 出し この 近く に ある か

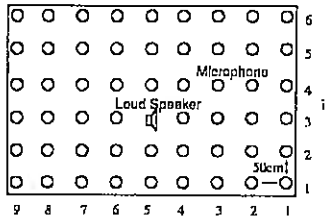
分散マイクロホン信号処理

- 空間ダイバーシティ
- 対数スペクトル重回帰と環境適応

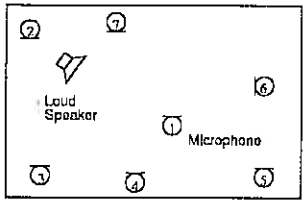
遠隔音声モデルと分散マイクロホン受音



(1)



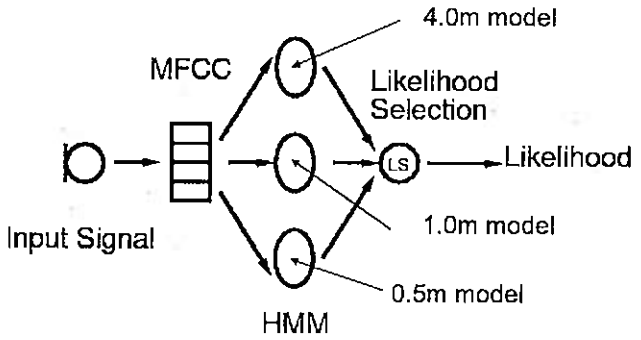
(2)



(3)

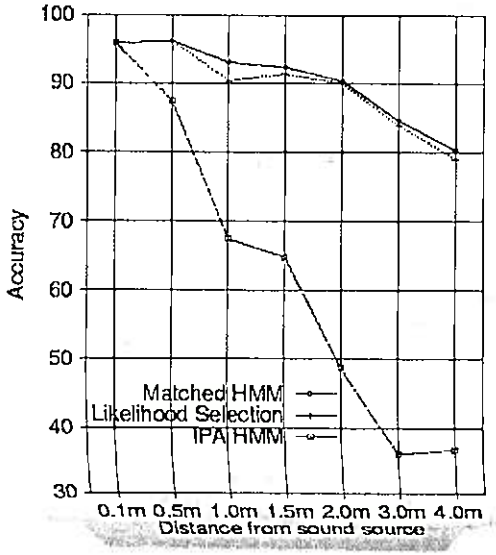
- 5.67m X 3.67m
- RT:320 ms

遠隔音声の最大尤度選択



1. Input speech is independently recognized by two or more models
2. Result of the highest likelihood model is selected

Experimental Results



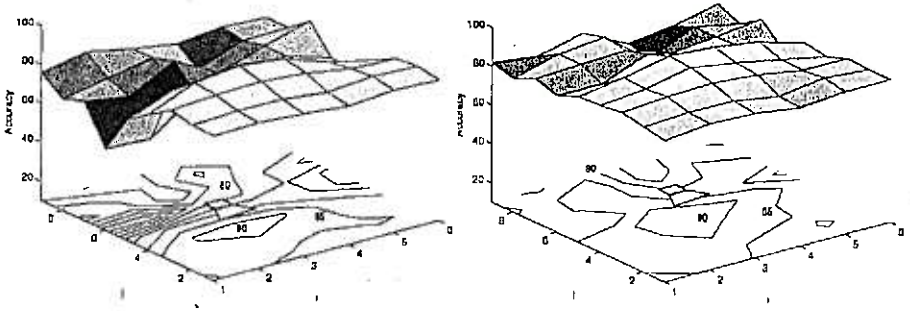
- One-third error rates compared with IPA HMM
- Likelihood Selection performs as well as Matched HMM

Maximum likelihood-based selection of the distant speech model is effective when the distance from sound source is unknown.

[Microphone Arrangement (1)]

Distribution of Recognition Rate

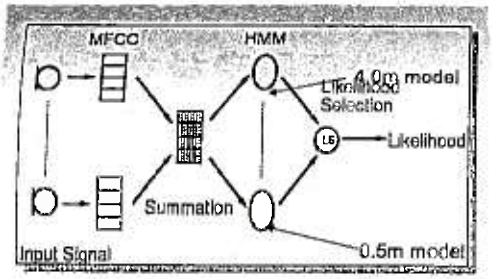
[Microphone Arrangement (2)]



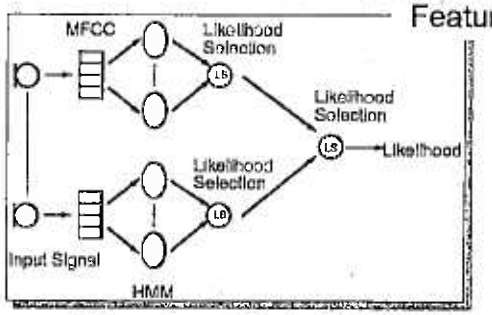
Only F-model (Average: 78.0%)

FSB-model (Average: 83.8%)

To attain high recognition rates, some acoustical variety of distant speech model is required.



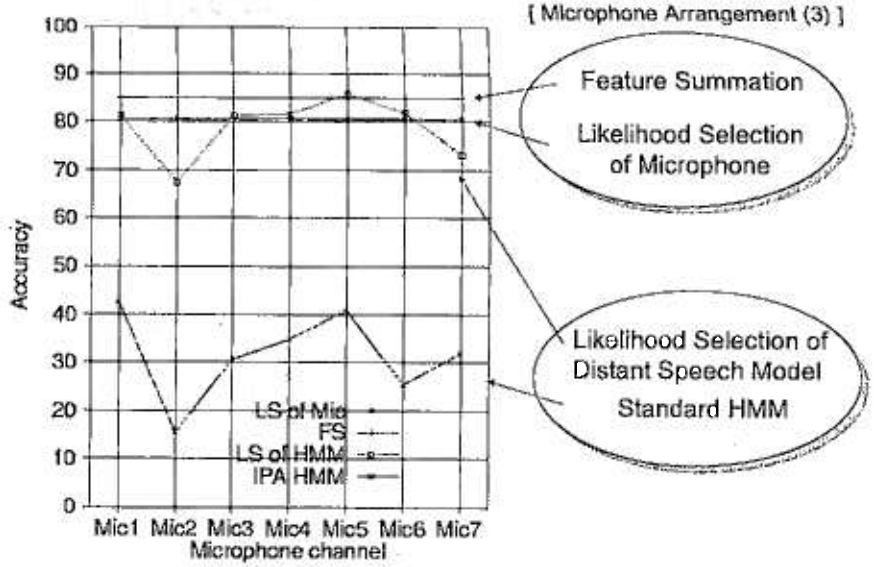
Feature Summation Method



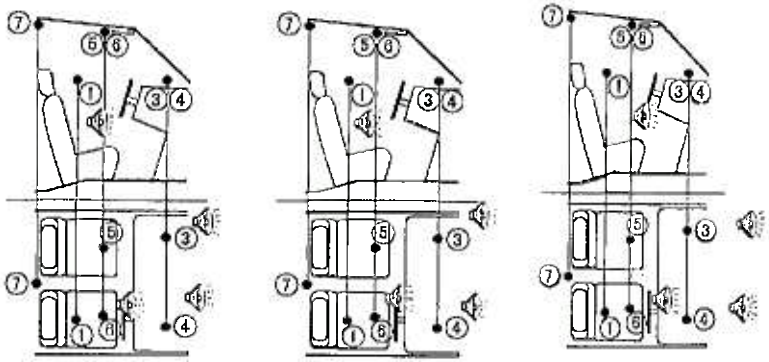
Microphone Selection

Experimental Results

[Microphone Arrangement (3)]

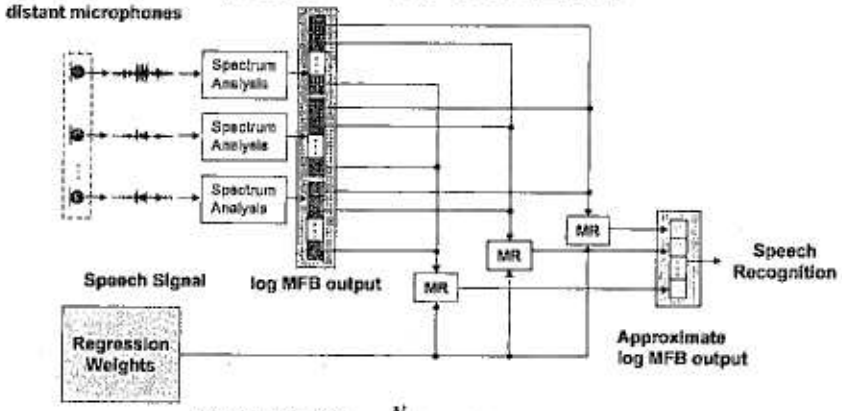


分散マイクロホンによる車内音声認識



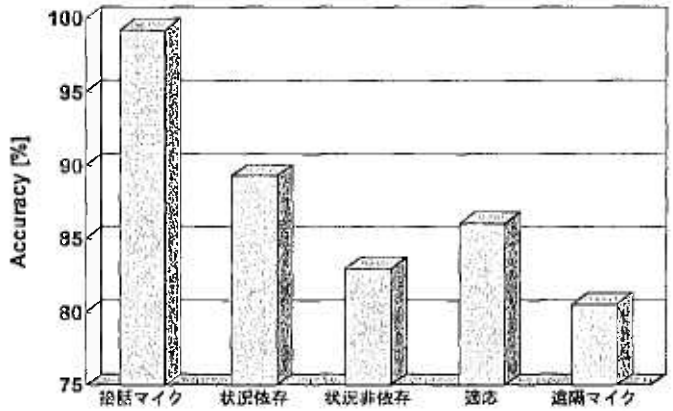
アイドリング　市街地走行　高速走行

分散マイクロホンを用いた遠隔音声認識（対数スペクトル重回帰）



$$\log \hat{X}_0(k) = \sum_{i=1}^{N} \hat{w}_i(k) \log X_i(k)$$

分散マイクロホンの性能



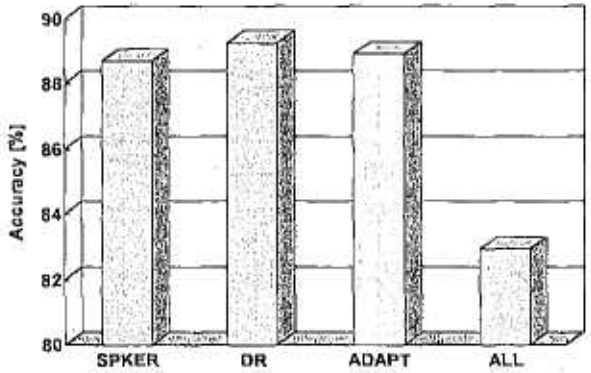
回帰重みの適応

- 分散マイクロホン毎のスペクトルパタンにより車内音環境を分類

分類結果

	normal	CD	fan lo	fan hi	window open
Class 1	2224	190	329	8	372
Class 2	440	2477	13	4	4
Class 3	25	20	2354	2684	35
Class 4	11	13	5	0	2289

回帰重みの適応の効果



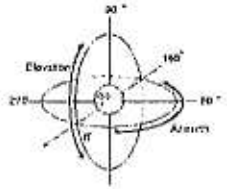
音響信号と空間物理

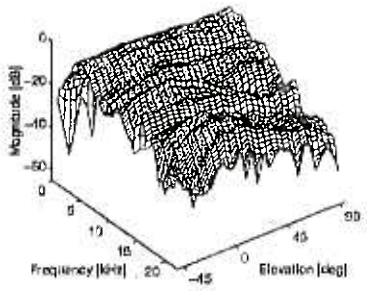
- 頭部伝達特性を用いた立体音響再生
- 信号分離

頭部伝達特性による立体音場再生

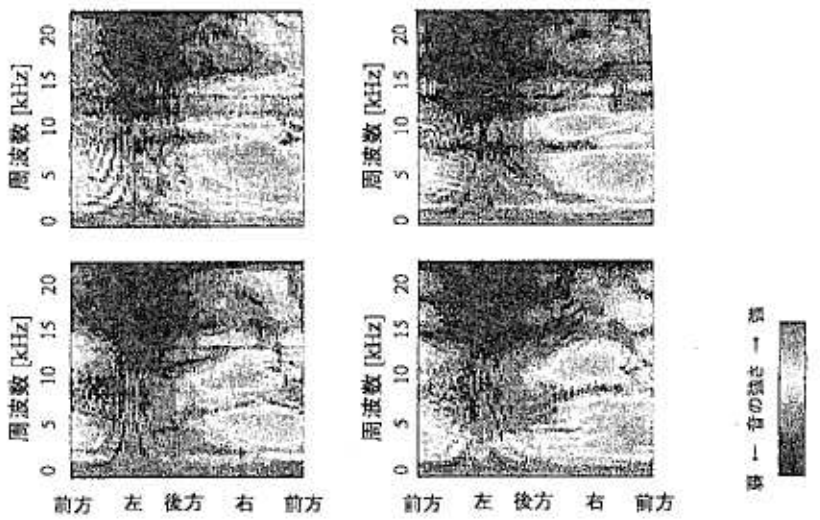
- 頭部音響伝達特性(HRTF)
  - 頭部・耳介形状に応じ音響伝達特性は個人到来方向で異なる
  - 頭部音響伝達特性を付与することにより、音を空間的に知覚させることが可能(音響バーチャルリアリティ)

$$H(z)=H(z;\theta,\varphi)$$



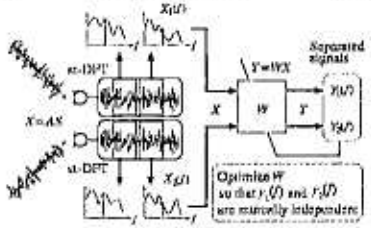


頭部音響伝達特性(HRTF)の特質



空間的アプローチと情報論的アプローチの統合

- 周波数領域でのブラインド信号分離
  - 周波数チャネルに分離信号に任意性が存在
- 空間指向特性の類似性に着目した入れ替え
  - 分離行列の重みを空間指向特性として解釈
- 分離性能
  - (1)混合音声　(2)分離音声　(3)分離音声



実世界データベースと競争的技術評価

- 車内行動信号処理に関する国際コンソシアム
- 車内行動信号コーパス

分散音響処理に関する国際コンペティション

- 国際ワークショップ
  - International workshop on DSP in Mobile and Vehicular Systems (April 3,4 2003, Nagoya)
  - Panel discussion:
    - Modelater: Sadaoki Furui (Tokyo Institute of Tech.)
    - Fumitada Itakura (Nagoya U.)
    - Huseyin Abut (Nagoya U./SDSU)
    - Morhan Sondhi (Bell lab.)
    - John Hansen (Colorado U.)

A proposal for

International Alliance for Advanced Studies on In-Car Human Behavioral Signals

Center for Integrated Acoustic Information Research

Research Focus

- Challenge
  - How we can built signal processing technologies for various human behavioral signals that are implicitly associated with human intention and the environmental realities?

Research Focus

- Sample Applications
  - Predictive driving interface
  - Personalization of car
  - Robust Communication for Safer Driving
- Fundamental Technologies
  - Speech/image signal processing
  - Interactive systems
  - Improved navigational systems

Research Roadmap

- In-car robust ASR
- AV integration for interface
- Driver Identification & Applications
- Integration & Interfacing with Navigational Systems for Improved Safety
- Modeling Driving Behavior
- Field Tests of Emerging Systems
- Recommendations to Scientific Community and Industry

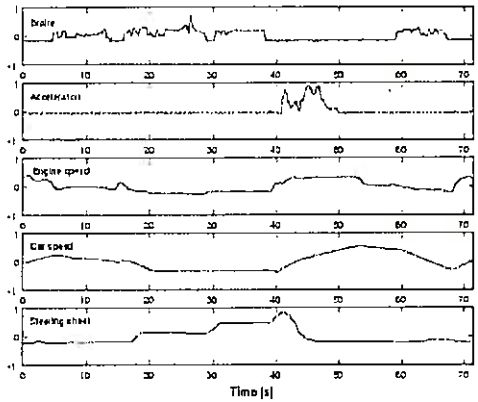
Corpus

- CIAIR Corpus of in-car human behavioral signals
  - Large amount
    - More than 800 speakers are involved
  - Real driving condition
    - Subjects are driving on a public street while making dialogues
  - Multi-mode dialogues
    - Dialogues with 1) a human navigator, 2) a WOZ system and 3) an ASR system are recorded.
  - Multi-media recordings
    - Recorded data include multi-channel audio, multi-channel video, vehicle-related information (speed, pedals, steering handle etc..), location.

International cooperation

- Corpus development and Sharing
  - Extending IHBS corpus
    - Multi-language
    - Extending driving conditions
  - Competitive evaluation of fundamental technologies
    - Word recognition using multiple microphones
    - Speech detection using audio and video signals
    - Driver recognition using driving signals
- Workshop

運転動作データ



brake, acceleration, engine speed, car speed, steering wheel

対話例（オペレータとの対話）

対話例（対話システム）

概要

- 多点（多重）観測と知的統合
  - 空間補間、情報統合
- 分散マイクロホン音声処理
  - ダイバーシティ、統合、適応
- 空間物理音響
  - 空間音響再生
- 実世界データベースと競争的評価
  - A proposal for International Alliance for Advanced Studies on In-Car Human Behavioral Signals