資料４

音声ヒューマンインターフェースの方向性

2003年3月14日

奈良先端科学技術大学院大学
情報科学研究科　教授

鹿野　清宏

総合科学技術会議重点分野推進戦略専門調査会
情報推進研究開発推進プロジェクトチーム
第4回会合

音声言語関連ヒューマンインターフェース

(1)　言語と情報発信
　言語の壁、　英語圏と非英語圏の格差の拡大

(2)　音声インタフェース利用の拡大
　キーボードから音声入力へ
　対話システム
　顔情報の利用
　無音声認識

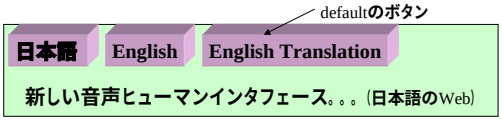
(3)　より自然な音声インターフェース技術
　マイクロホンアレーによるハンズフリー
　音場制御による自由なバージインシステム

言語と情報発信

Webの作成　　世界への情報発信、ITビジネス

日本語のWeb
&
英語のWeb

作成に時間がかかる
アップデートが大変



日本語のWebには、自動英語翻訳モードを自動的に！

機械翻訳の研究をもう一度

Web自動翻訳研究

機械翻訳

日本の得意分野
(1) 多くの研究者、研究ポテンシャル
(2) 不十分ながら各種ソフトウェア
(3) 大規模対訳データの存在
(4) 統計的手法の発展

LDC　米国
ELRA　EU
GSK?　日本

問題点
情報の取得・発信のための言語処理基盤
著作権、フリーソフトウェア、分野ごとの辞書などの未整備、言語処理技術、
コーパスベースの言語知識獲得技術

解決の方策
(1) 国策として、データの利用の許諾と管理組織の設置
(2) 大学を中心とした言語研究者の結集
(3) オープンソースのフリー翻訳ソフトウェア、辞書の開発

音声インタフェース利用の拡大

音声認識
　高精度のオープンソースプログラムやツール
音声合成
　高品質かつ多様な音声合成
応用：
(1) ボイスポータル (電話からの問い合わせ)
　日本では？　i-modeの影響
(2) 携帯、PDA　　苦戦？
(3) カーナビ　　普及して有望、多言語への拡張？
(4) 放送やディクテーション　　？？？
(5) 聴覚障害者などの補助　　国策が必要
(6) ディジタルデバイドの解消　？？
(7) 語学学習　　？？
(8) IT教育　気軽に音声でWeb検索

近未来の生活におけるITの予測

- Networking：世界中がつながった無線やブロードバンドの普及
- Robots：　生活の中でのロボットの存在
  音声認識/合成, 顔表情認識による対話機能
- Wearable：より小さく高機能化された個人用携帯端末
  音声認識/合成、無音声認識、無音声電話

音声インタフェース利用の拡大

奈良先端大での研究開発例

(1) 音声対話システム
ロボットとの対話、
エージェントとの対話 (公共の場での利用) 、
視線などの顔情報の利用

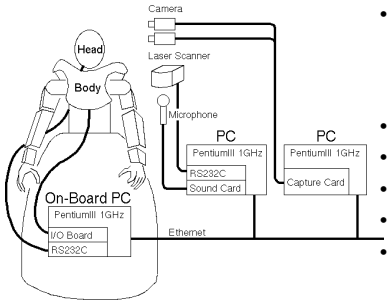
(2)無音声認識&無音声電話
新しい音声メディアの創造

人とロボットの対話システム (Receptionist Robot: ASKA)

- 研究科の受付に置いた案内ロボット
- 実環境におけるマルチモーダル対話実験システム
- 顔をみつけてユーザの方向を向き、音声を認識して、音声とジェスチャーで応答するロボット

2万語の大語彙連続実時間音声認識

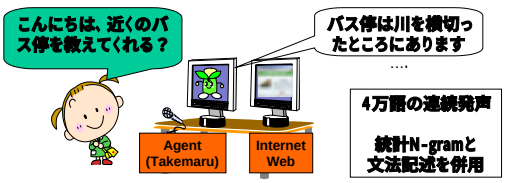
ハードウェア構成 (Hardware configuration)



- 基本ハードウェア
  本体: Tmsuk IV
  頭: Infanoid Robot
  両眼
- レーザスキャナー
- マイクロフォン
- スピーカ
- CCD カメラ
- Linux PC

ASKA: Video

たけまる音声ガイダンスシステム (Takemaru Speech Guidance System)



4万語の連続発声

統計N-gramと文法記述を併用

生駒市北コミュニティセンターに昨年11月から常設
音声でエージェントに話しかけると、応答音声とWebの表示で応答
子供などがWeb検索などのIT技術に親しむ, IT教育？

一般の人に使われる音声認識ガイダンスシステム
(1)頑健な音声認識システムのフィールドテスト
(2) マンマシン対話の枠組みでの対話音声データの収集
(3) 話者環境適応手法の評価

画像処理を用いた顔情報計測技術

従来の視線計測装置（頭部装着型）

開発した視線計測装置（非接触型）

小型カメラ＋ノートパソコン

計測精度

顔の位置：2mm
姿勢：2度
視線方向：5度

処理速度

30Hz～80Hz
（カメラによる）

応用範囲

- コンピュータ・インターフェース（ex. 視線マウス）
- ロボット・インターフェース（ex. 操作対象の指示）
- 人間工学（ex. 人間に負担の少ないデザイン）
- 安全システム（ex. 運転手の居眠り検出）
- 心理学（ex. 視覚認知実験）
- 福祉機器（ex. 視線による車いすの操縦）

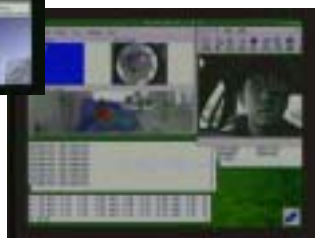
画像処理を用いた顔情報計測技術

顔情報の計測

計測結果（顔の動き，視線の動き）をそのまま使ってエージェントの顔を動かした様子

ドライバ計測への応用

ドライバの注視行動を道路シーンや他のセンサ情報とともに記録し，行動解析に用いるシステムを構築



画像処理を用いた顔情報計測技術

対話ロボットへの応用

ロボットは人間に見られていることを認識できる → アイコンタクトしながらコミュニケーションできるロボットを実現

知的車いすへの応用

レバーを使わなくても見ている方向へ自動的に進んでくれる電動車いすを実現

無音声認識　(NAM)
Non-Audible Murmur Recognition

音声認識・合成の利用範囲を拡大
無音声電話？？

Yoshitaka Nakajima@NAIST

つぶやき声を観測する方法
Analyzing Method for Non-Audible Murmur (NAM)

- A man is now murmuring something, but you cannot hear.
- What would you do to examine what content is spoken?
- Are you a engineer, a computer scientist, a biologist, or… a ? Maybe…..

A MURMURING MAN

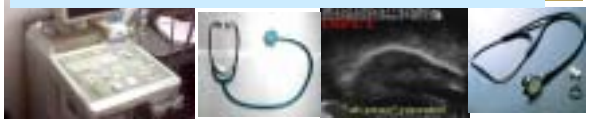
Dr. Nakajima is not a physicist but a physician.
中島さんは、医者です！！

つぶやき声の観測方法
Three methods for NAM recognition

A) **聴診器**を体表に当てて、発話器官により調音された呼気音を聞き取る. AUSCULTATION

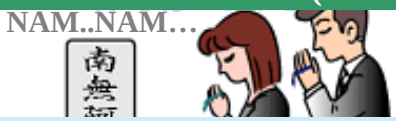
B) **低周波・超音波**の発振器を体表に当て、体表の別部位から受信して、声道や軟部組織の共振フィルタ特性を解析する. PERCUSSION

C) **超音波イメージング装置**のプローブを顎下に当て、舌と唇の動きを読む. ULTRASONOGRAPHY



NAMの定義 (Definition of NAM)

Non-Audible Murmur (NAM)



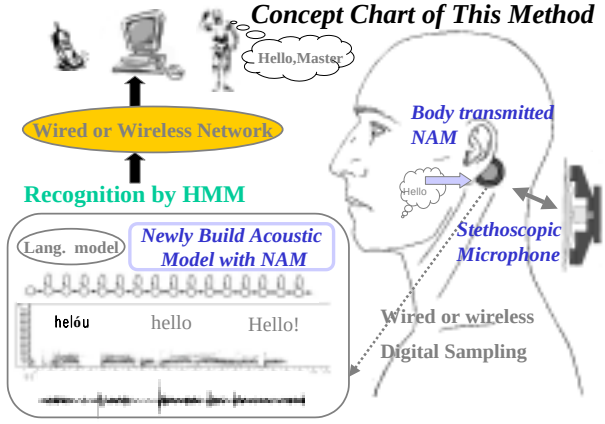
- 周囲の人に内容を聴取不能な口の中での発話行動
- 発話器官のフィルタ特性により調音された声帯振動を伴わない軟部組織伝達の無声呼気音
- ささやき声→外部限定聴者を想定
- 口パク→呼気を伴わない

無音声認識は何がうれしいか

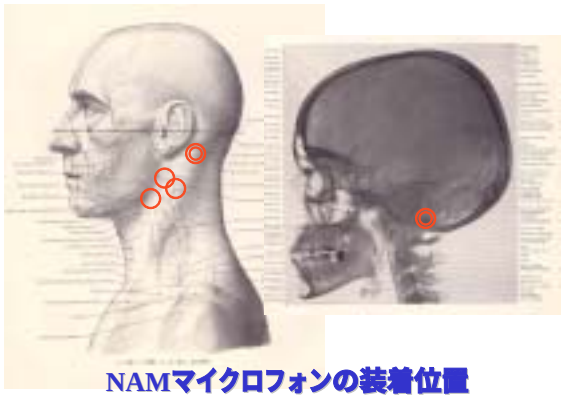
音声認識の便利さに加えて

- 空中音声を相手にしないので雑音に強い
- 周囲に騒音を撒き散らさない公共性
- **音声言語のパブリシティーコントロール**が可能
- 声を出さないので恥ずかしくないし疲れない
- 不特定話者を想定せず個人に特化されたモデルを作ればよい
- 人類の新たな音声言語文化を創り出す可能性
  (無音声電話…)

Concept Chart of This Method



Personal Digital Assistant Terminal



NAMマイクロフォンの装着位置

Location of the NAM Microphone Attachment

NAM音声 (Comparison of Three Speeches)

NAM音声の大語彙連続音声認識実験

Evaluation of NAM Speech Recognition

20k Dictation Task, Speaker-dependent Monophone model from ATR 525 phoneme balance sentence utterances × 4 and 1258 ×2 , Julius decoder

発声環境	発声文数	単語正解率	単語認識精度	置換誤り率	脱落誤り率	挿入誤り率	誤り率
1.Quiet	24	**93.6**	93.3	4.7	1.7	0.3	6.7
2.Music	24	**91.1**	90.0	6.7	2.2	1.1	10.0
3.TV	24	**89.7**	89.2	9.2	1.1	0.6	10.8
平均/合計	72	**91.5**	90.8	6.9	1.7	0.7	9.2

1. In a quiet room
2. With background music at the volume we usually enjoy (a Bach Concert for Violin and Orchestra)
3. With the sound of TV news