SCOPE 若手ICT研究者等育成型研究開発（092108002）

発話障害者のコミュニケーション支援のための携帯電話用読唇システムの開発

研究代表者
齊藤 剛史　九州工業大学

研究期間　平成21年度～平成23年度

研究開発の目的

- 発話障害者がこれまで利用できなかった通話機能を利用可能な「携帯電話用読唇システム」の構築
  1. 使用するカメラや撮影方向、画像サイズなどの撮像機構の確立
  2. 認識対象とする発話内容の設定および発話シーンの撮影
  3. 撮影画像から口唇領域の自動抽出手法の提案
  4. 認識に有効な特徴量および認識手法の提案
  5. 処理の実時間化
  6. プロトタイプシステムの構築および被験者実験による定量的評価の実施

読唇の目的

- 音声認識の補助
  - 高騒音環境における認識精度の向上
- 聴覚・発話障害者のコミュニケーション支援
- 無音声認識
  - 公共の場所における音声通話の実現
  - 映像のみからの発話内容の復元

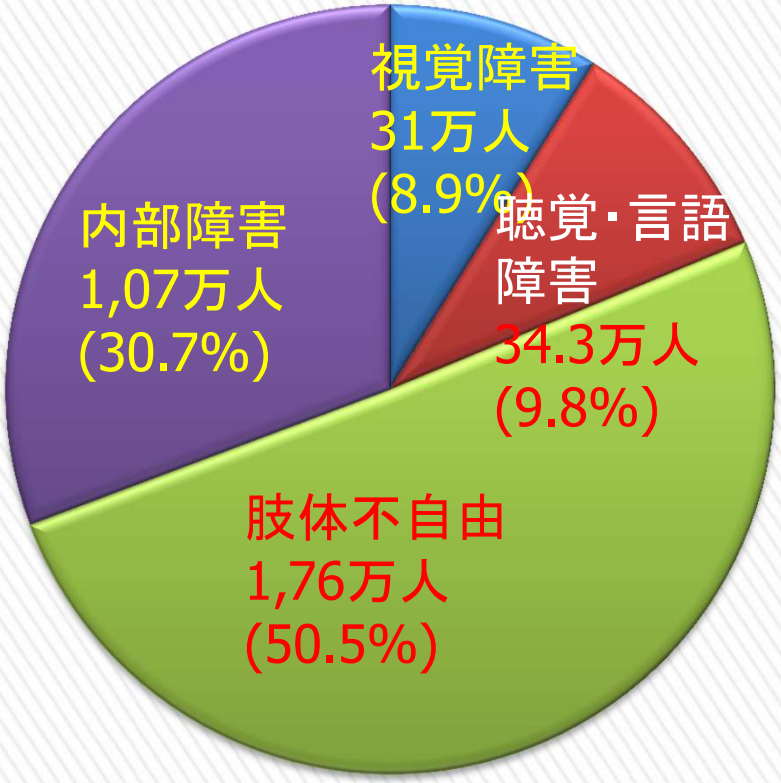
読唇？

- 人の読唇精度は**30～40%**
- 工学的な研究は**1980**年代～
- 読話、口話、読唇
  - 「2001年宇宙の旅」人工知能HAL9000は横顔の読唇

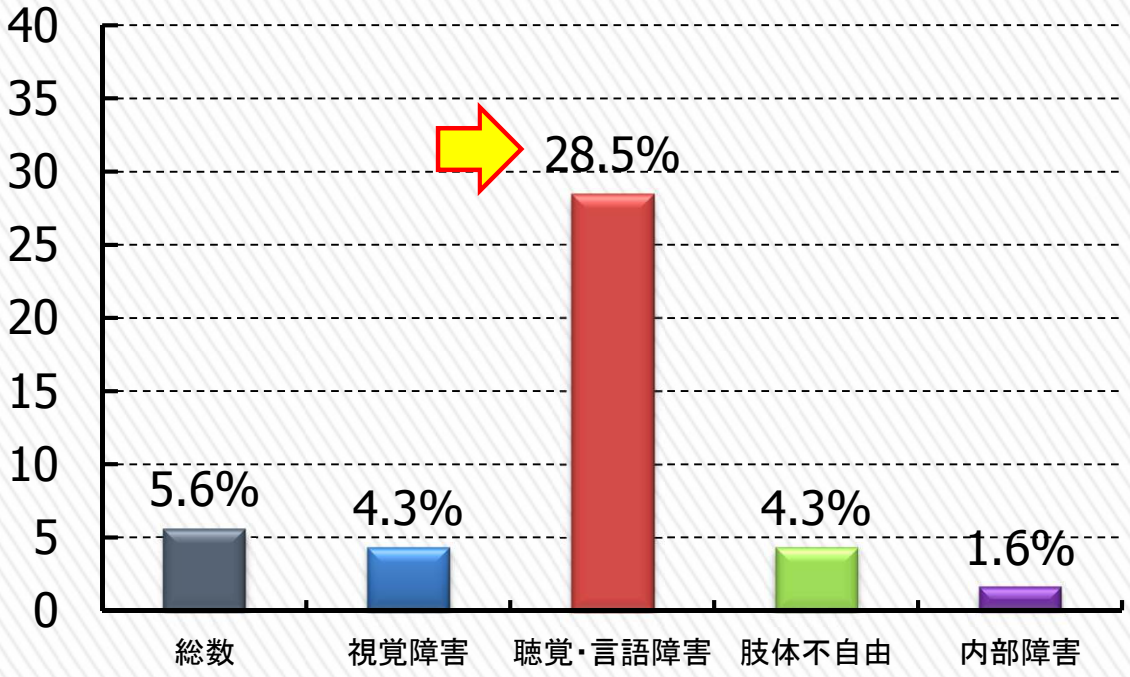
研究の背景

平成18年身体障害児・者実態調査結果

厚生労働省



障害の種類別にみた身体障害者数



外出するうえで、または外出しようとするうえで困ること
人と話をすることが困難

研究の背景

■ 障害者のコミュニケーション手段

- 残存聴力
  - 補聴器
  - 人工内耳
- 視覚情報
  - 手話、指文字
  - 読話
  - 筆談

研究の背景

■ 発話障害者

e.g. 喉頭摘出、気管切開

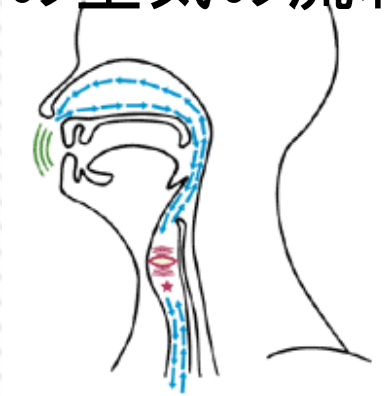
- コミュニケーション支援システムの開発は大切な課題
- 音声情報のみを利用した音声認識（automatic speech recognition; ASR）システムは利用できない。

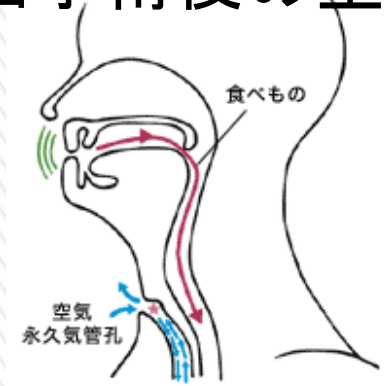
- 急性期の治療や気管切開を受け意識は回復しているが発声できない患者は常時**4～5**人（1地方病院の人数）
- 全国の病院数8670施設（平成22年10月）を考慮すると**3～4万**人の患者
- 口頭摘出による失声者数は**2～3万**人（赤木、日本医療機械学会大会2008）。

従来技術とその問題点

■ 喉頭摘出者の発声法

- 通常の空気の流れと喉頭摘出手術後の空気の流れ





- 声を発声する三つの方法

食道発声

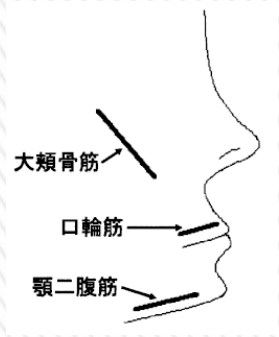
シャント発声法

電気式人工喉頭

従来技術とその問題点

- **音声**　非接触デバイス、8kHz、16kHz
  - 実用（カーナビ、Siriなど）
- **肉伝導**　接触デバイス、16kHz
  - "非可聴つぶやき認識"、中島他、信学論（2004）
  - 人に聞こえない、体表から直接センシング、外部雑音に対して頑健

- **筋電信号**　接触デバイス、2kHz
  - "無発声音声認識：筋電信号を用いた声を伴わない日本語5母音の認識"、眞鍋他、信学論（2005）
- **読唇**　非接触デバイス、30Hz



従来技術とその問題点

- 読唇
  - Audio-visual ASR（Automatic Speech Recognition）
    - “Evaluation of real-time audio-visual speech recognition”
      - Shen et al. (Proc. of AVSP2010)
    - “Real-time lip reading system for isolated Korean word recognition”
      - Shin et al. (Pattern Recognition 2011)
  - 口形＋キー操作
    - “MouthType: Text entry by hand and mouth”
      - Lyons et al. (Proc. of CHI2004)
  - Visual-only ASR
    - “A novel transducer: From lip motion to voice message”
      - Saitoh et al. (Proc. of MVA2009)

新技術の特徴・従来技術との比較

■ カメラの利用

- 非接触型装置
- 空気の流れがなくても可能

■ シンプルな構成

- PCとカメラ

■ 無音声認識

- 周囲の人に迷惑を与えない
- 静かな場所で利用できる
- 声を出せない人が音声コミュニケーションできる。

期間内の実施項目

- プレ評価実験の実施
- 発話障害者およびそのコミュニケーション支援に関する調査
- 読唇システムにおける撮像機構・手段の確立
- 認識対象の発話内容の決定および発話シーンの撮影
- 撮影画像から自動的に口唇領域を抽出する手法の提案
- 口唇の動きと認識精度に関する検討
- 発話シーンのフレームレートと認識精度の検討
- 読唇に有効な視点の検討
- 単語ベース読唇と文章ベース読唇の検討
- プロトタイプシステムの開発および評価実験
- 研究成果の公開
- 読唇に有効な顔モデルの検討

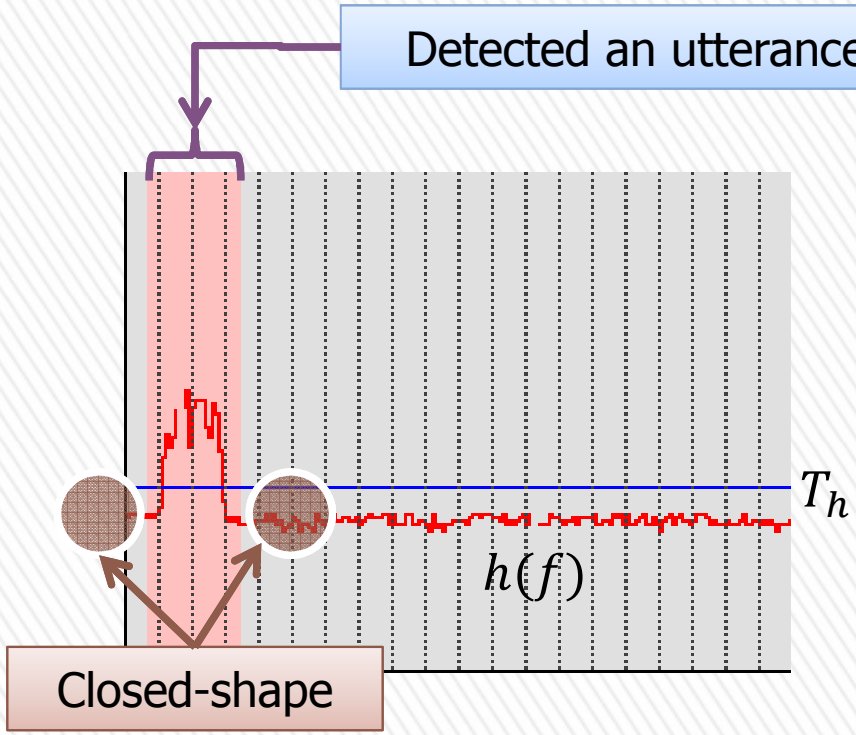
読唇アルゴリズム

読唇アルゴリズム

- 口唇領域の抽出
  - Viola-Jones顔検出器
  - Active Appearance Model（AAM）を用いた顔と口唇の抽出
- 発話区間の自動検出
  - 口の開閉を利用
- 特徴量
  - Appearanceパラメータ（AAM）
- 認識手法
  - DPマッチング

発話区間の自動検出

- 口唇の高さ$h(f)$を利用
  - 閉唇口形：$h(f) \leq T_h$
  - 発話区間：二つの閉唇口形に挟まれた区間
  - シンプルだが有用なアプローチ



コミュニケーション支援システム

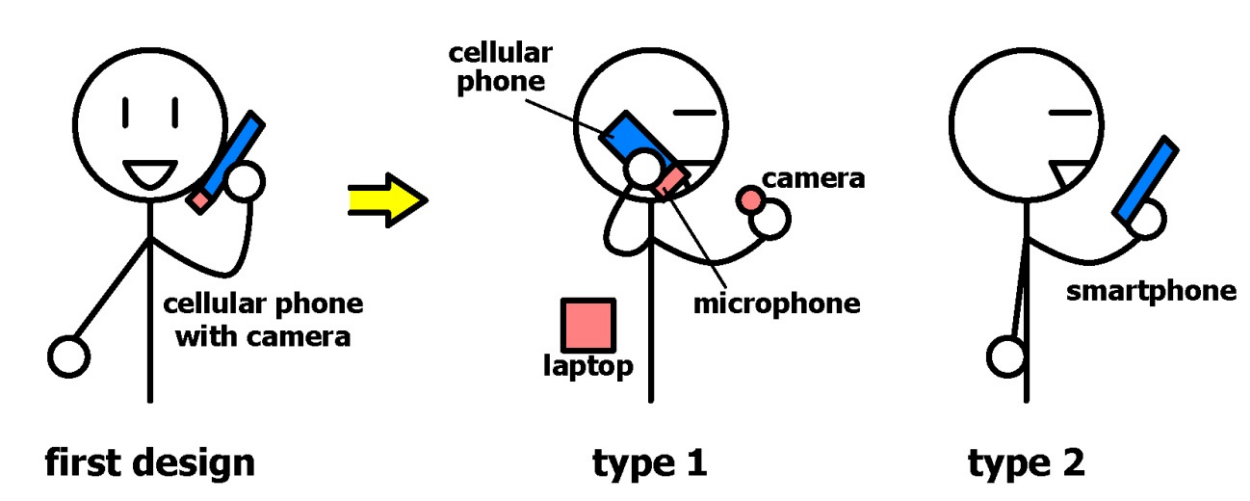
プロトタイプシステムの開発

システム要件

- 発話内容は単語単位あるいは短文単位とする。
  - 単語ベース読唇と文章ベース読唇の検討結果に基づく
- 正面顔画像を用いる。
  - 読唇に有効な視点の検討結果に基づく
- 抽出処理時間は1フレームあたり333ms以下（10fps以上）とする。
  - 発話シーンのフレームレートと認識精度の検討結果に基づく
- 認識処理だけでなく学習データ登録作業も手軽に操作できるようにする。
  - 被験者が容易にシステムを操作可能にする

システム要件

- ノートPCベースのシステムを構築する
- 発話終了後2秒以内に認識結果を出力する。
- 認識結果をテキストとして画面に表示するだけでなく音声メッセージも出力する。
- 操作の利便性を考慮して、把持しやすいワイヤレスコントローラを利用する。

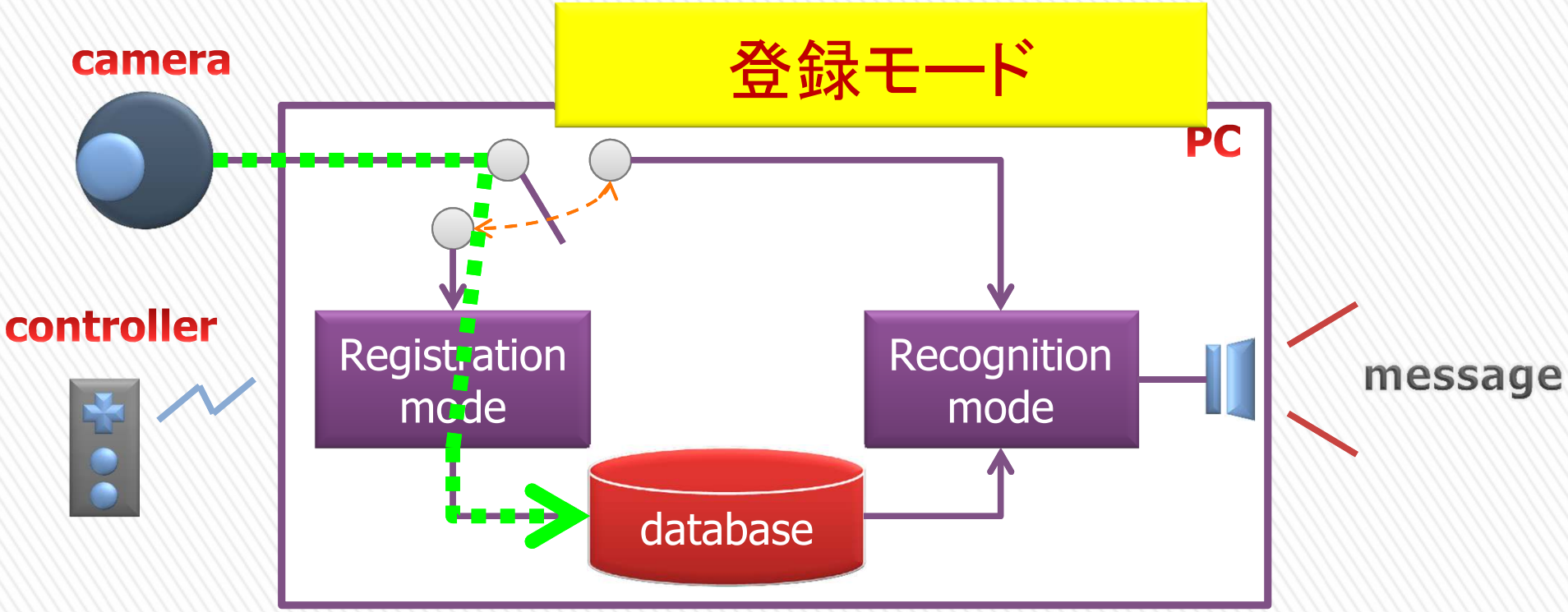


システム機能

- 読唇機能
- 発話区間の自動検出機能
- 目標文入力のための誤認識文の退避機能
- メッセージを一文ずつ伝達するだけでなく、複数の単語を組み合わせたメッセージを伝える2種の伝達機能
- 光源環境の影響を軽減するためのカメラ制御機能

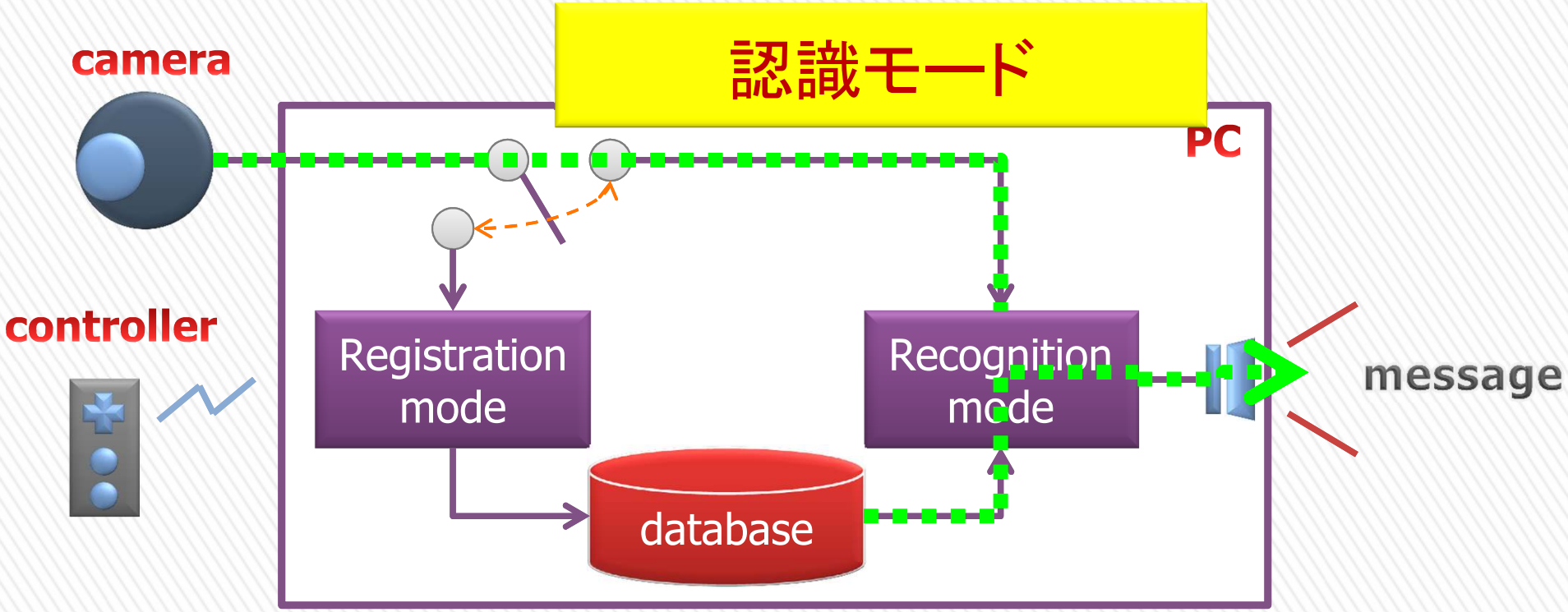
コミュニケーション支援システム

- 2モード
  - 登録モード
  - 認識モード
- ハードウェア：PC、カメラ、コントローラ



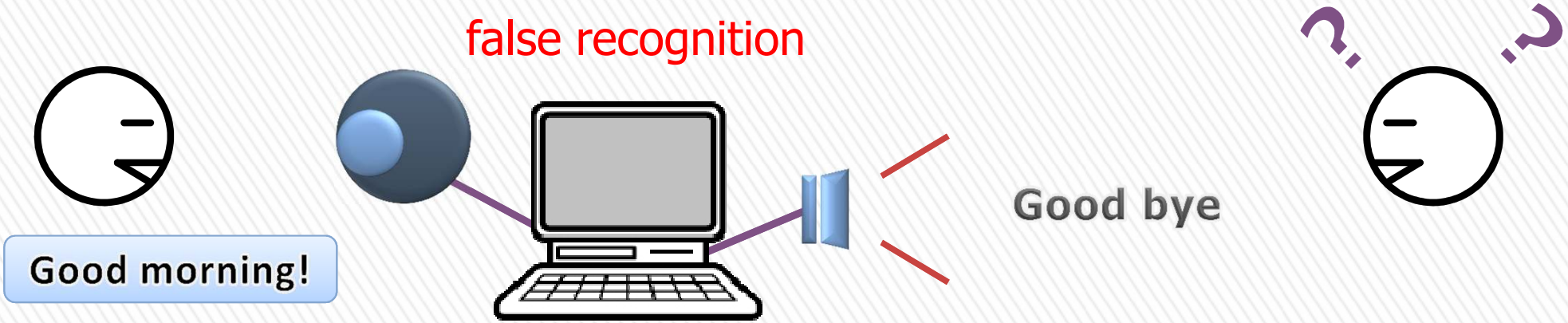
コミュニケーション支援システム

- 2モード
  - 登録モード
  - 認識モード
- ハードウェア：PC、カメラ、コントローラ



コミュニケーション支援システム

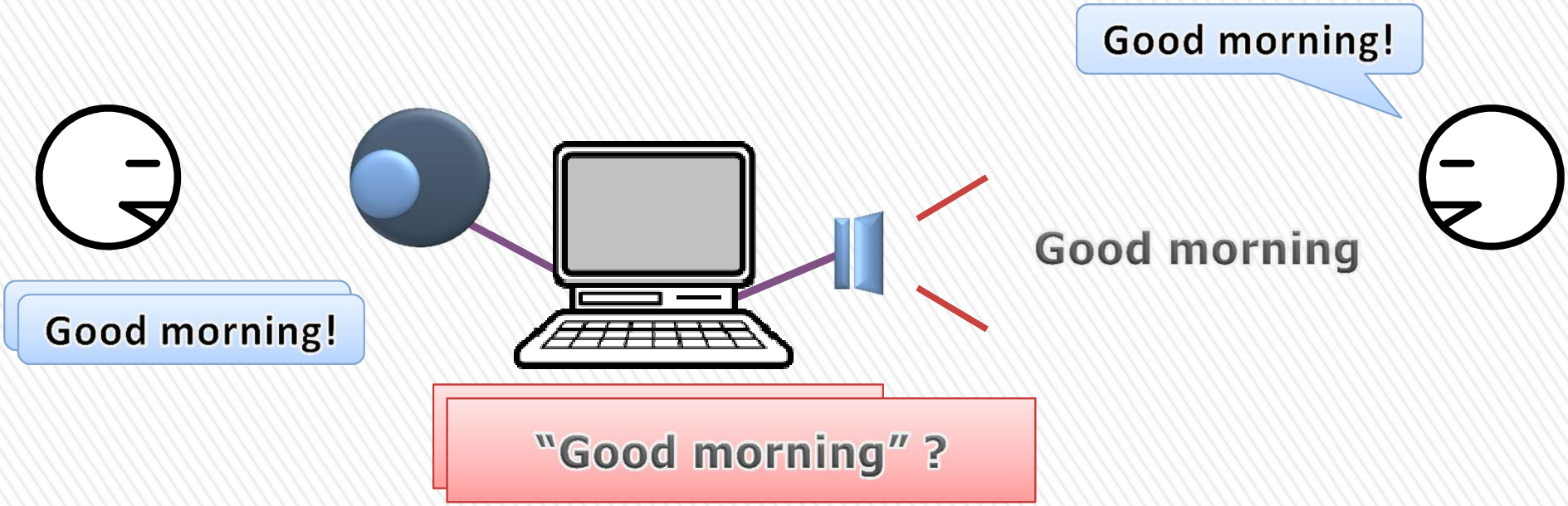
- 誤認識の場合、認識後に認識結果をそのまま出力すると・・・



話し相手とコミュニケーションに失敗する

コミュニケーション支援システム

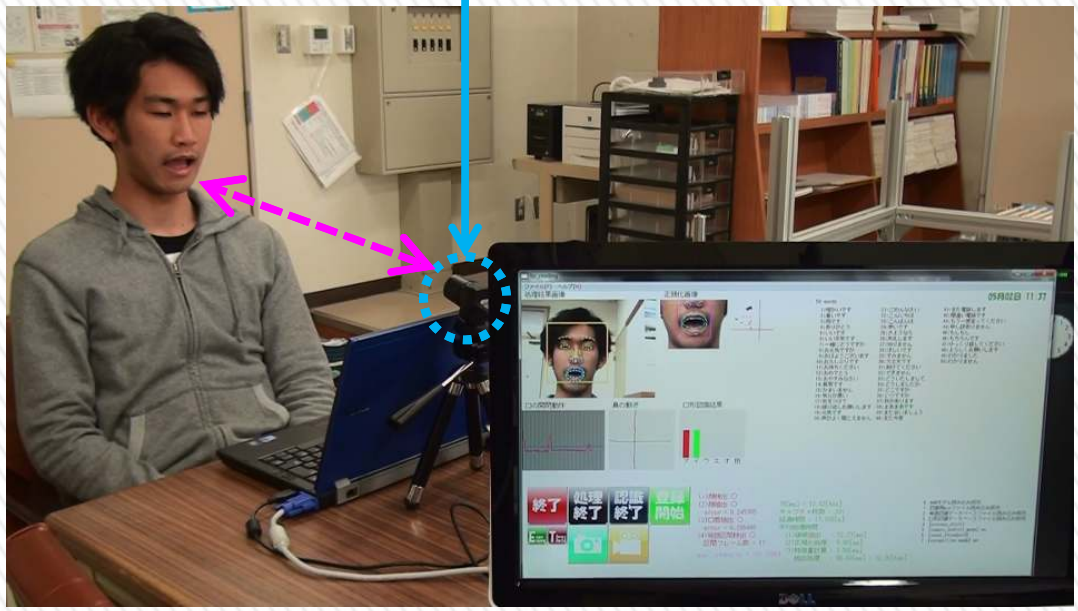
- 誤認識による誤メッセージの伝達を防ぐ
  - 認識結果をユーザに表示し、認識結果が正しければ話し相手にメッセージを伝える。



⇨ 話し相手と正確なコミュニケーションを実現

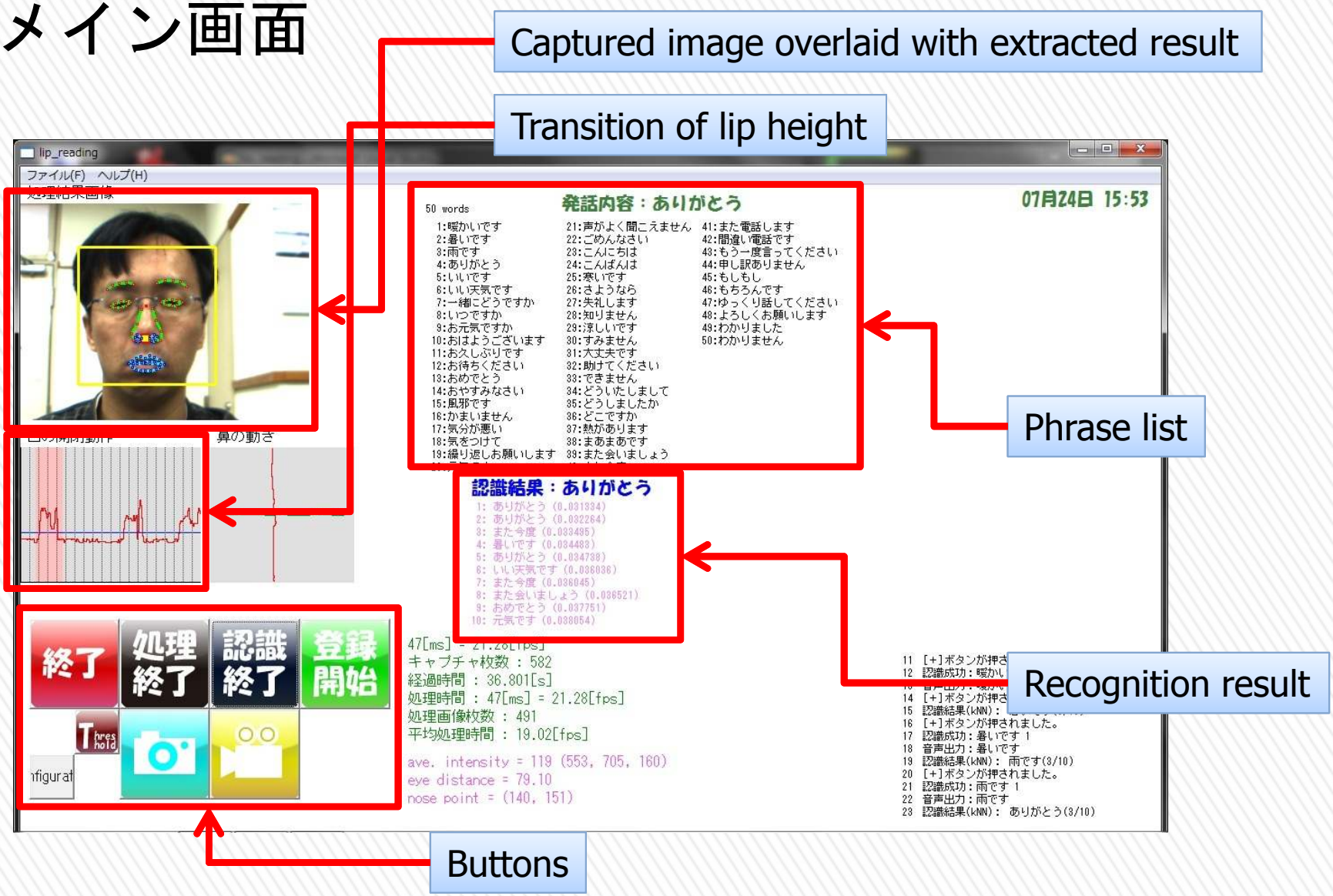
プロトタイプシステム

- PC: Laptop (CPU: Intel Core2 i5-520M, 2.40GHz)
- Camera: Point Grey Research, Chameleon
- Lens: Fujifilm, YF4A-2
- 画像サイズ：**320×240**画素
- 顔ーカメラ間の距離：**60**cm
- 処理時間: **22.3**fps



プロトタイプシステム

- メイン画面



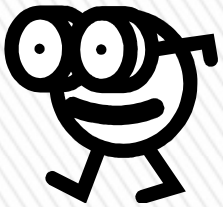
評価実験

実験条件

- 認識対象語：
  - 電話会話50文
- サンプル数：
  - 1文につき10サンプル
- 被験者：
  - 4名（男性、健常者）
- 姿勢：
  - 座位

#	発話内容	#	発話内容
1	暖かいです	26	さようなら
2	暑いです	27	失礼します
3	雨です	28	知りません
4	ありがとう	29	涼しいです
5	いいです	30	すみません
6	いい天気です	31	大丈夫です
7	一緒にどうですか	32	助けてください
8	いつですか	33	できません
9	お元気ですか	34	どういたしまして
10	おはようございます	35	どうしましたか
11	お久しぶりです	36	どこですか
12	お待ちください	37	熱があります
13	おめでとう	38	まあまあです
14	おやすみなさい	39	また会いましょう
15	風邪です	40	また今度
16	かまいません	41	また電話します
17	気分が悪い	42	間違い電話です
18	気をつけて	43	もう一度言ってください
19	繰り返しお願いします	44	申し訳ありません
20	元気です	45	もしもし
21	声がよく聞こえません	46	もちろんです
22	ごめんなさい	47	ゆっくり話してください
23	こんにちは	48	よろしくお願いします
24	こんばんは	49	わかりました
25	寒いです	50	わかりません

モデル構築



- Speaker B
  - 顔モデル用（10枚）

  - 口唇モデル用（22枚）

登録枚数	Speaker				Ave.
	A	B	C	D	
顔モデル	19	10	8	15	13.0
口唇モデル	19	22	29	25	23.8

実験結果

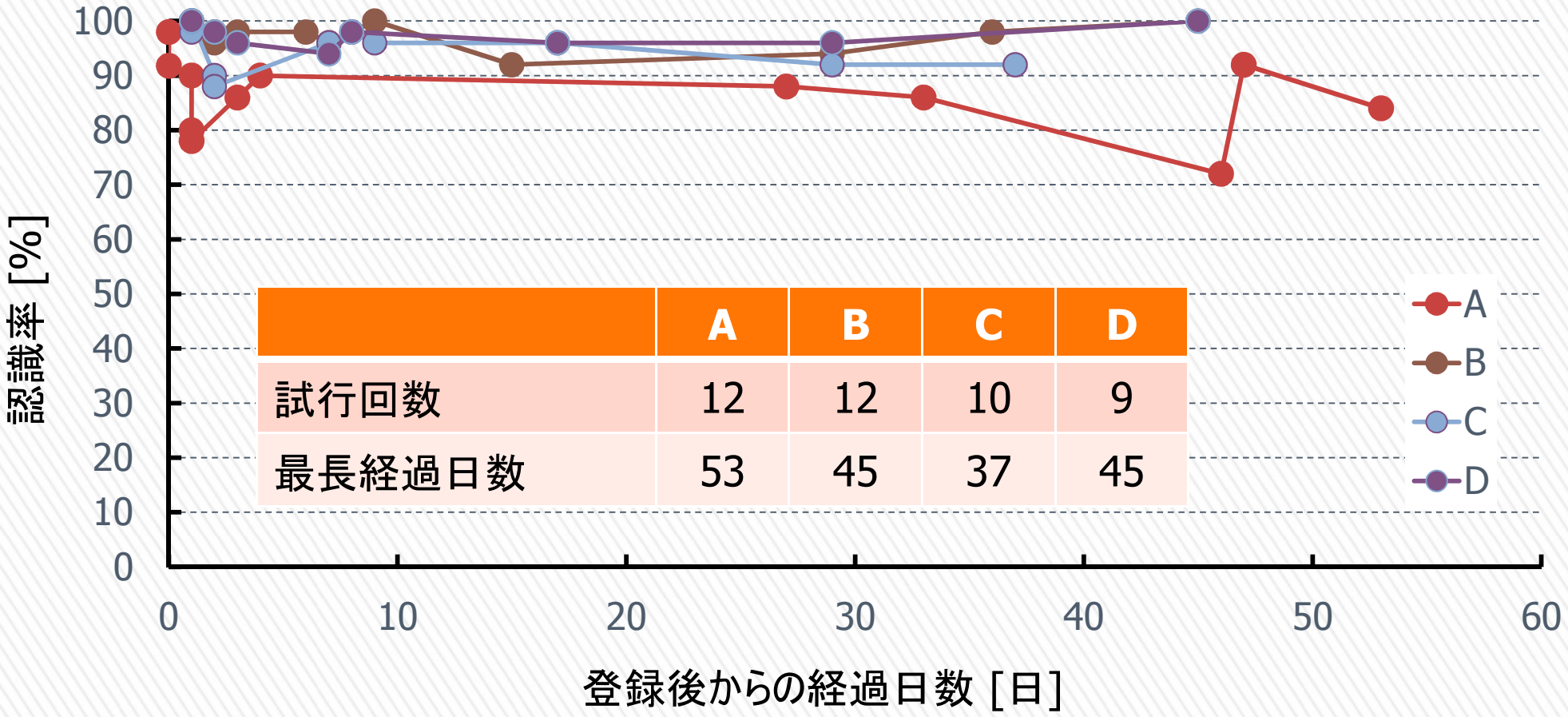
- N_{frame} ：1文あたりの平均登録フレーム数
- $R[\%]$ ：平均認識率
- t_s ：認識時間
- t_v ：認識後から発話開始までの時間

	Speaker				Ave.
N_{frame}					50.6
$R[\%]$					94.4
$t_r[\mathrm{s}]$	0.112	0.230	0.280	0.184	0.198
$t_v[\mathrm{s}]$	1.27	1.25	1.52	1.59	1.41

- 平均認識率：**94.4**%
- 認識時間：**0.2**秒

実験結果

■ 各被験者の認識率の推移



	A	B	C	D
試行回数	12	12	10	9
最長経過日数	53	45	37	45

研究開発の結果及び成果

- 基礎研究の実施
  - システム開発のための撮影機構の確立
  - 読唇に有効な視点
  - フレームレートの検討
  - 口唇領域の自動抽出手法など
- 完成度の高いプロトタイプシステムの開発
  - 実時間読唇の実装や登録モードや認識モードなど
- 日本語50文に対して被験者4人の評価実験
  - 2ヶ月弱の試行
  - 平均認識率94%
  - 発話終了後2秒以内に音声メッセージを出力

研究成果の社会的意義
社会への波及効果

- 本研究成果はコミュニケーション支援システムとして障害者の生活の質の向上に貢献する。

- 本システムは声を発する必要がなく、発話障害者だけでなく、騒音時や公共の場所などで声を発することが望まれない場所でも利用可能であり、音声認識インタフェースとしての利用が期待されている。