人と人との会話を解き明かす

コミュニケーションを科学する 映像音声技術

NTTコミュニケーション科学基礎研究所 メディア情報研究部 大塚和弘

講演の流れ

- ・なぜ,「会話」なのか?
- ・実時間会話シーン分析システム
 - 全方位センサによる会話の観察
 - 画像系技術
 - 音響系技術
- ・会話構造の推定
- ・未来
 - アプリケーション
 - 技術課題など

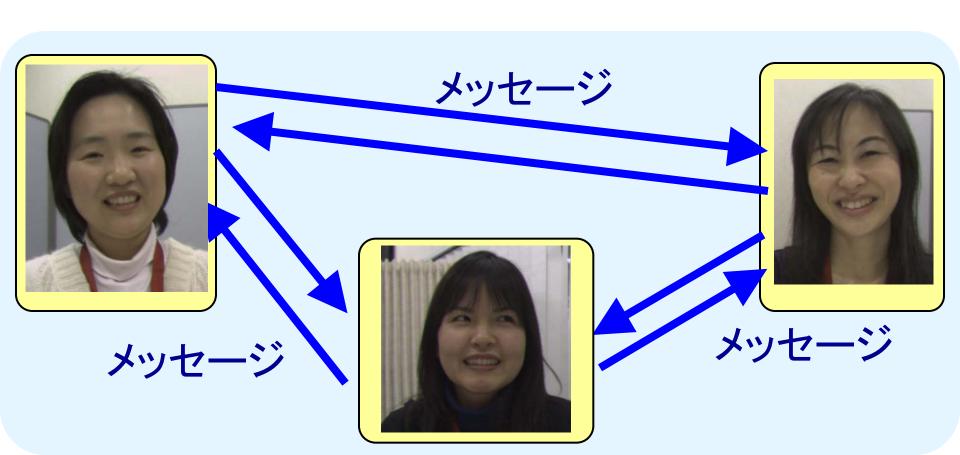
なぜ、「会話」なのか?



コミュニケーションの障壁を克服したい。

研究の目的

- コミュニケーション科学
 - -メッセージを分析し、人間同士の 自然なコミュニケーションを自動的に理解する



非言語行動・非言語メッセージ

対面コミュニケーションにおいて重要な役割を持つ

発 話 音響的 韻 律 情報 視 線 頭部ジェスチャ 顔の表情 視覚的 情報 手のジェスチャ 姿勢

クロスモーダル・インタラクション

送り手の行動 受け手の行動 発 話 発 話 韻 律 韻 律 視 線 視線 頭部ジェスチャ 頭部ジェスチャ 顔の表情 顔の表情 手のジェスチャ 手のジェスチャ 姿勢 姿勢

複数のモダリティー間のインタラクションが鍵を握る

会話シーン分析の問題設定

(非)言語行動を画像·音声信号として観測し、 そこから会話の6W1Hを自動的に認識する

when
いつ?where
どこで?who
誰が?whom
誰と?what
何を?how
どのように?why

- 例題
- ・誰が誰に向かって話し掛けているか? who-to-whom
- ・誰が誰に対してどのように反応しているか?who-whom-how
- ・誰が何を話しているのか? who-what
- ・何故,彼/彼女は怒っているのか? who-how-why

実時間会話シーン分析システム

2008.5~

デモシステムのターゲット

発 話

・誰がいつ話しているか?

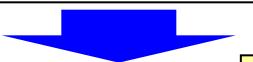
→ 話者ダイアリゼーション(Speaker Diarization)の問題



視線

・誰がいつ誰(どこ)を見 ているか?

→視覚的注意の焦点(Visual Focus of Attention)

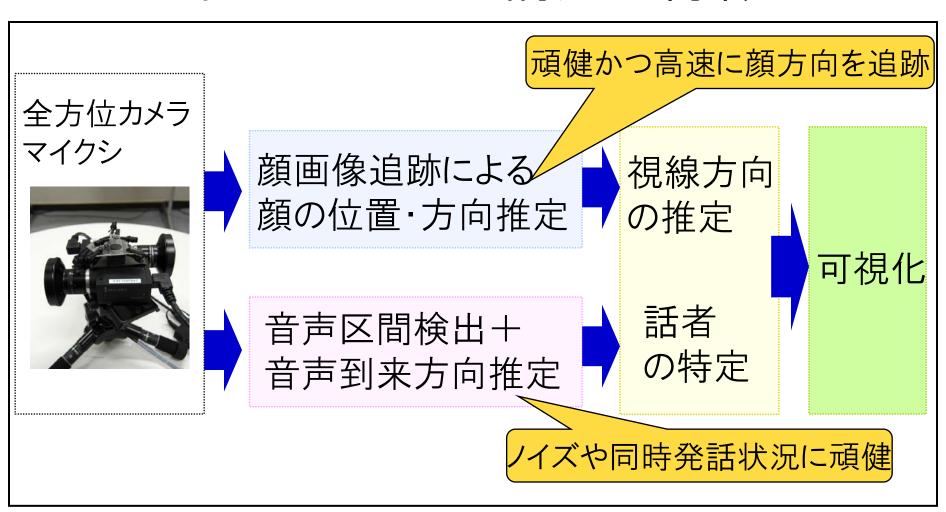


「いつ誰が誰に向かって話しているか?」「いつ誰が注目を集めているか?」

whenいつ?

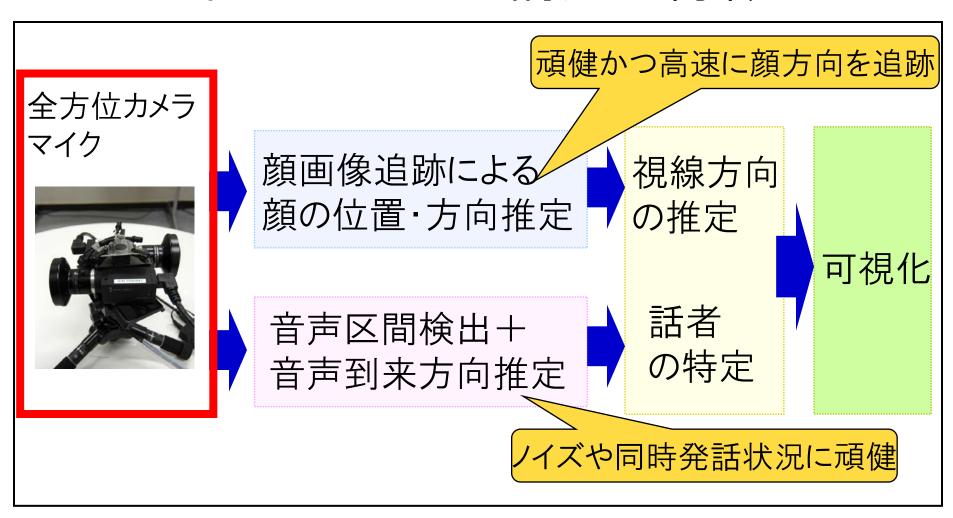
who 誰が? whom 誰と?

本システムの構成と特徴



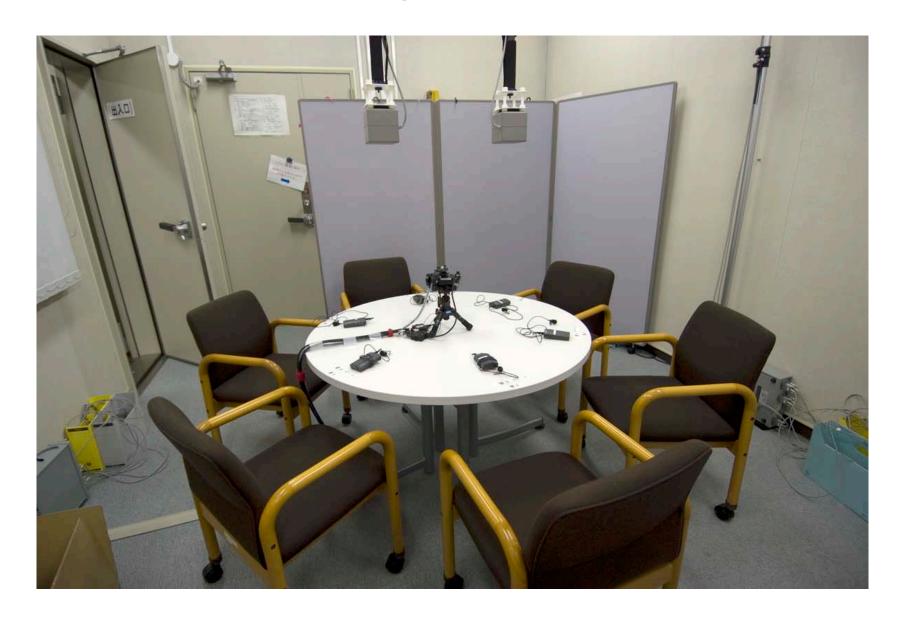
リアルタイム動作する 多人数マルチモーダル会話シーン分析システムは世界初!!

本システムの構成と特徴



リアルタイム動作する 多人数マルチモーダル会話シーン分析システムは世界初!!

ミーティングルーム



ミーティングシーン

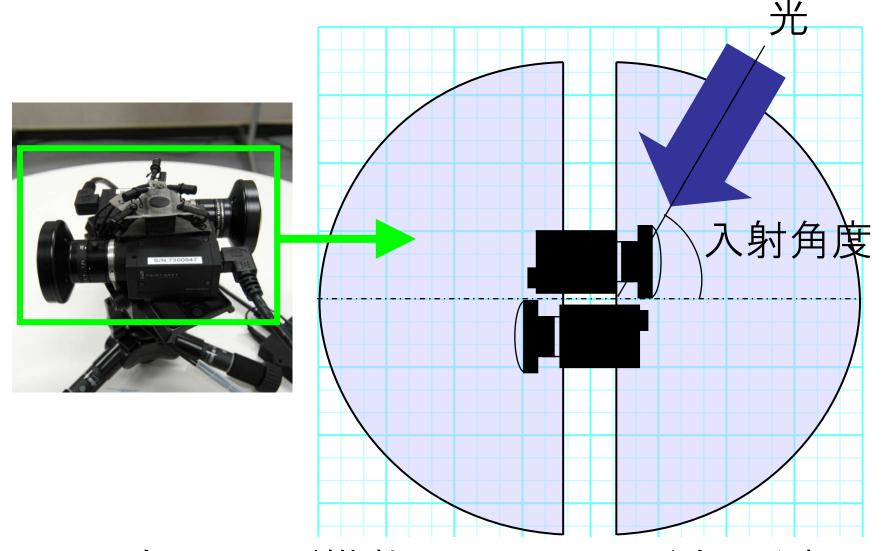


全方位カメラマイク





全方位カメラ



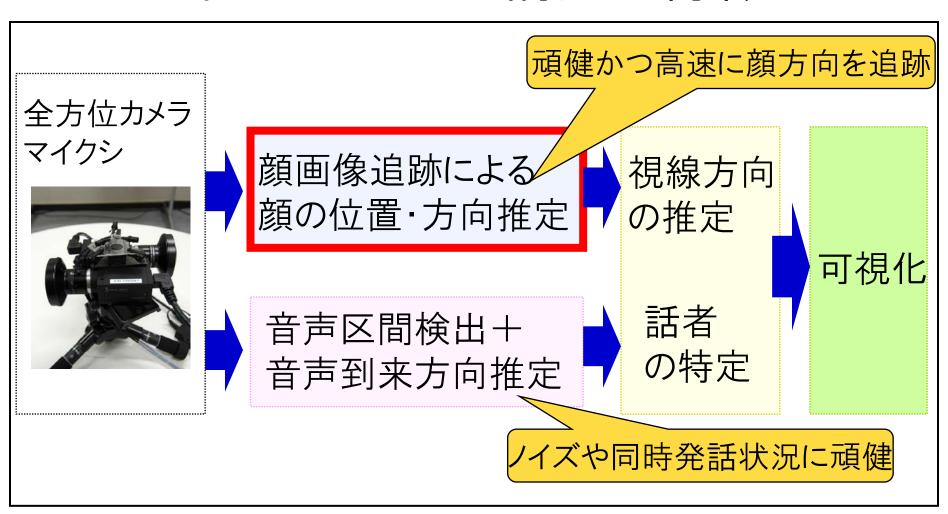
2つの魚眼レンズ搭載カメラによりほぼ全周を撮影

マイクロフォンアレー

3本のマイクによりマイクアレーを構成 音 120度/ 3本のピンマイク

デモ:全方位カメラによる撮影画像とその表示方法

本システムの構成と特徴

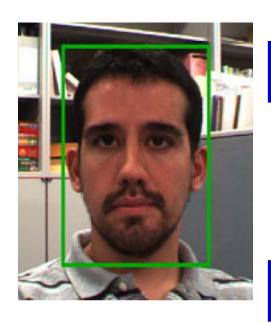


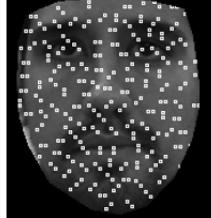
リアルタイム動作する 多人数マルチモーダル会話シーン分析システムは世界初!! デモ: 顔方向追跡

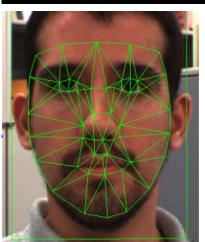
顔方向追跡:初期化

疎テンプレート追跡法

顔領域の検出

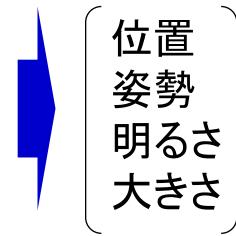






注目点集合

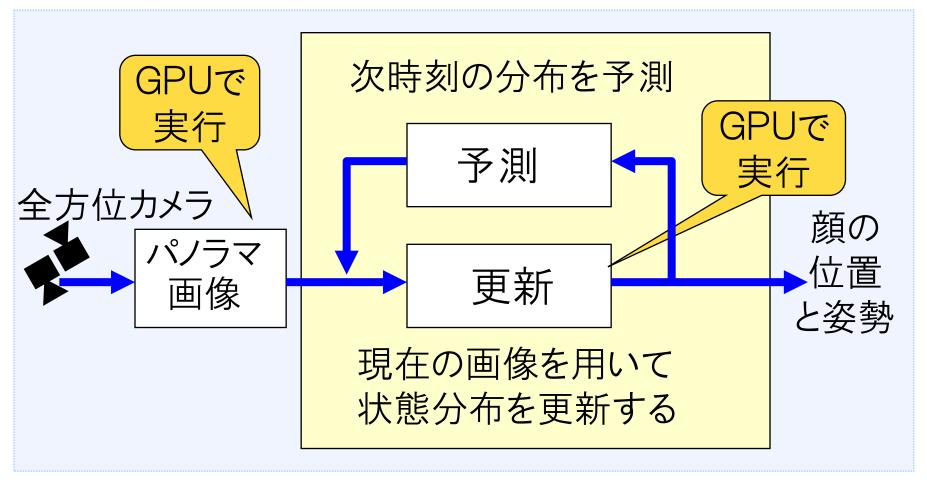
顔テンプレートの状態



形状モデル

顔方向追跡:追跡

顔テンプレートの状態(位置と姿勢)の分布を パーティクルフィルタを用いて求める



メモリーベース・パーティクルフィルタ

新しいパーティクルフィルタの枠組みを提案

- ・過去の状態履歴を活用することで「予測」を向上
- ・効果1:複雑な動きへの対処
- し・効果2:追跡失敗からの復旧性能の向上

提案法:過去の長期的な動きの性質を利用











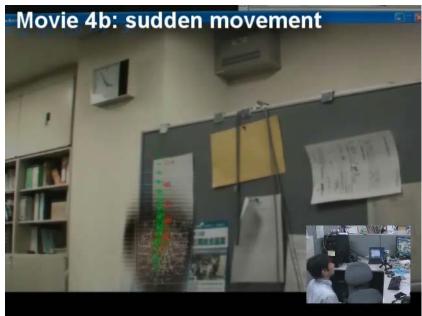


従来法:直近の情報のみ利用

Mikami, Otsuka, Yamato: "Memory-based particle filter for face pose tracking robust under complex dynamics", Proc. IEEE CVPR2009

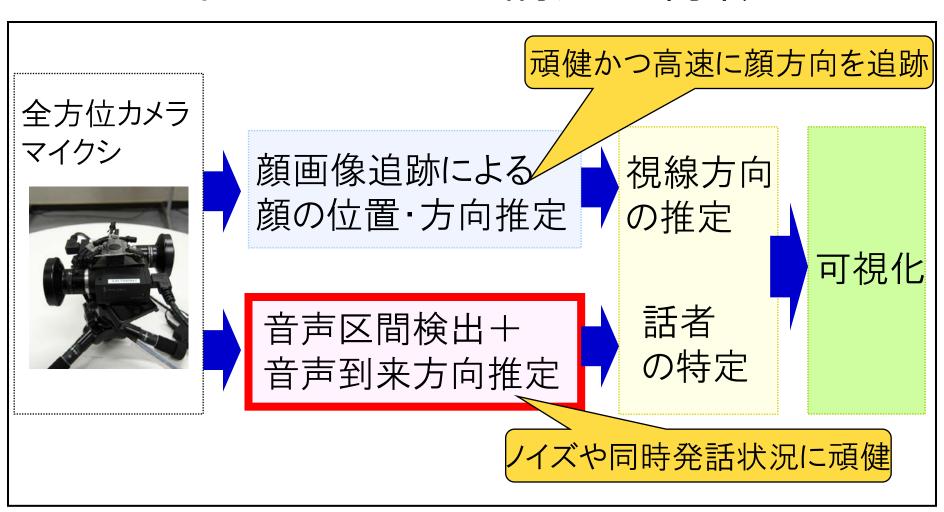
ムービー:M-PF







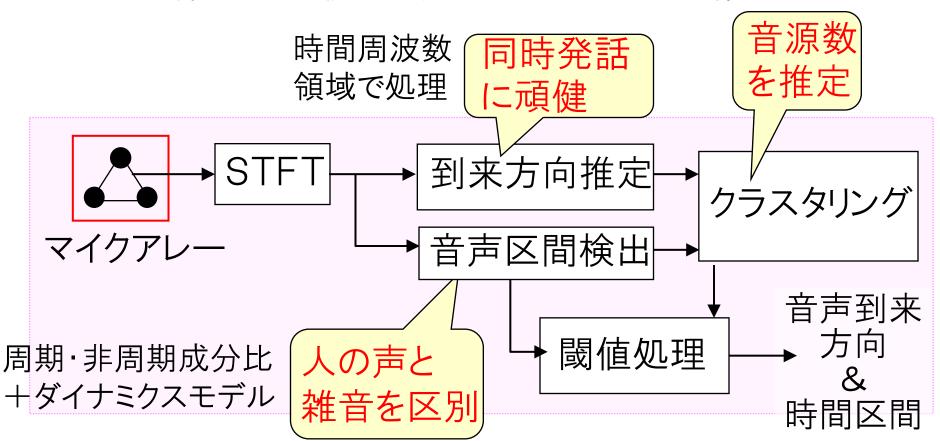
本システムの構成と特徴



リアルタイム動作する 多人数マルチモーダル会話シーン分析システムは世界初!!

音響信号処理

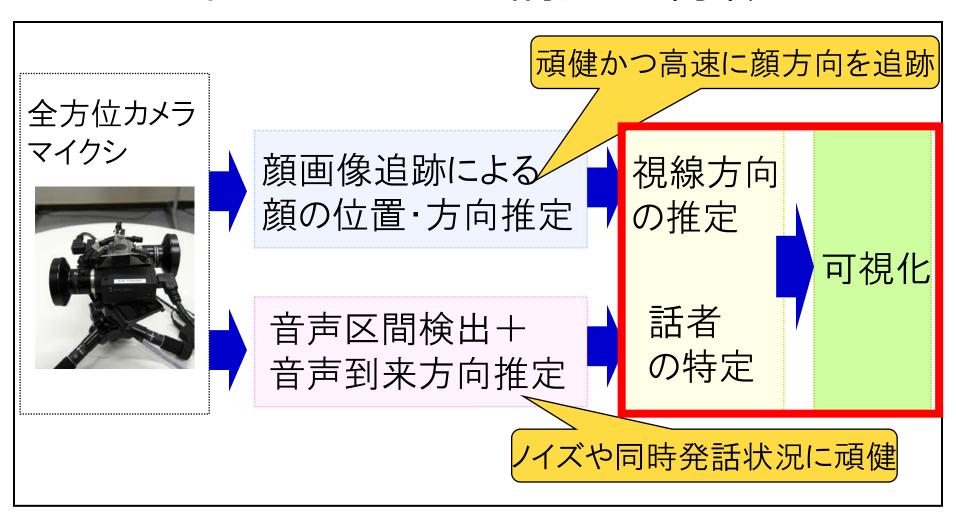
各人の発話区間の検出と声の到来方位角の推定



S. Araki, et al. A DOA based speaker diarization system for real meetings. In *Proc. HSCMA2008*, pages 29-32, 2008

デモ: 音声区間検出

本システムの構成と特徴



リアルタイム動作する 多人数マルチモーダル会話シーン分析システムは世界初!! デモ:可視化

ダイアグラム表示 音声強調

本システムの特徴(まとめ)

- ・全方位カメラ・マイクシステム
- ・画像上での高速・頑健な顔方向追跡
- ・頑健な音声区間検出
- ・リアルタイム性
- · 3次元可視化
- ・音声強調

when
いつ?where
どこで?who
誰が?whom
誰と?what
何を?how
どのように?why

会話構造の推定

when
いつ?where
どこで?who
誰が?what
誰と?how
どのように?why

会話モデルのコンセプト

推定の対象

会話の構造:参加者間でのメッセージ伝達のパターン

「誰が誰に話し掛けているか?」「誰が誰の話を聞いているか?」 "who is talking to whom?" "who is listening to whom?"

視線パターン (視線の方向)

インタラクション構造

「誰が誰を見ているか?」
"who is looking at whom?"

「誰が誰に反応するか?」
"who responds to whom?"

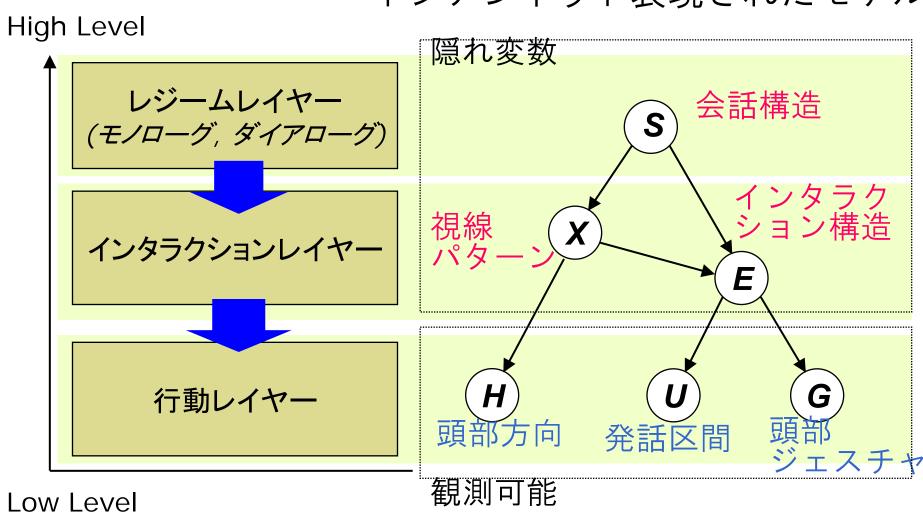
観測の対象

発話の有無 頭部方向 頭部ジェスチャ

会話構造と非言語行動との関係を確率的にモデル化

階層的モデル構造

ベイジアンネット表現されたモデル



頭部ジェスチャによるインタラクション

頭部ジェスチャは、話し手・聞き手、双方にとって重要な手掛かり

話し手の頭部ジェスチャ

話し掛け行動 -リズムとり -強調



問い掛け

明示的に他者からの 応答を要求する



聞き手の頭部ジェスチャ

傾聴行動 Back-channel response 相槌, 頷き



返答

明示的に態度を表明 e.g. 同意, 否定



話し掛け一傾聴

問い掛け一応答

推定結果の可視化例:空間的な表示



白い円:中心人物(話者など)を表す



未来へ

Future Directions

今後の方向性

- ・もつと高性能・頑健に
 - 顔追跡の頑健性向上,
 - 発話検出の精度向上, 残響除去
- ・もっとマルチモーダル
 - 顔表情や韻律から感情を推定
 - 視線の方向を直接計測
 - 音声認識による言語情報の利用
 - マルチモーダル統合による上位の会話の状態の推定
- ・もつと見せる/使える
 - アプリケーション開発

アプリケーション

- ・遠隔映像会議
 - リアルタイムの分析結果を用いた映像生成
- ・会議アーカイブ構築
 - 会話の状態をキーとした検索・閲覧
- ・心理学研究、心理カウンセリング、人物評価
 - 個人の特性の評価, グループの特性の定量化

動画像からの表情認識技術



顔向き変動に対してロバストな

表情認識手法を開発

Kumano, et al. "Pose-Invariant Facial Expression Recognition Using Variable-Intensity Templates", Int. Journal of Computer Vision, 2008

演技的な表情から会話中の微細な自然表情へ

会話の中の微笑・哄笑

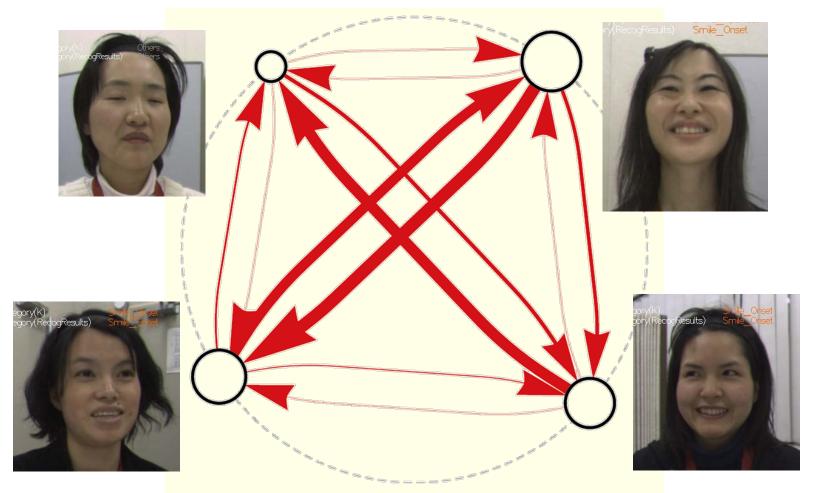


上段: 人間の判定

下段:認識結果

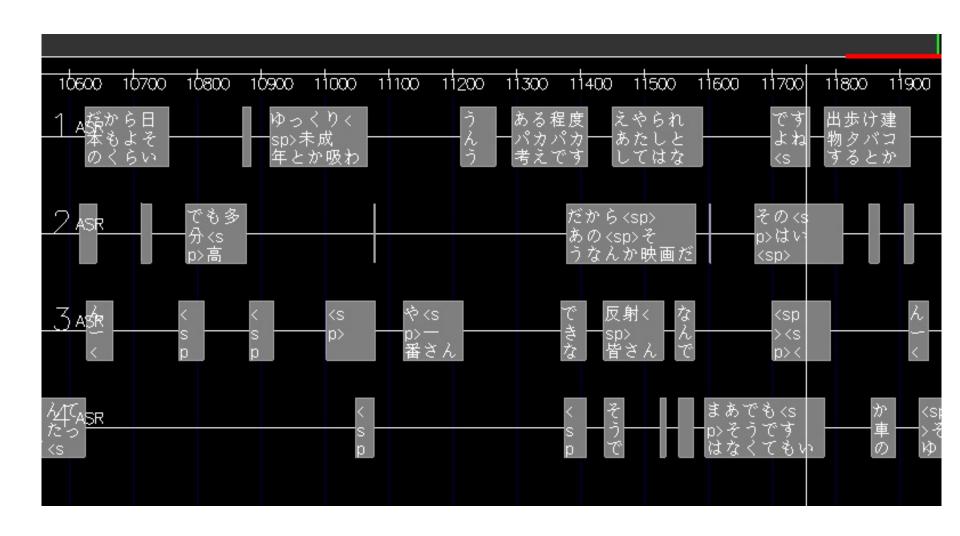
対人感情の推定

誰が誰にどれくらい微笑んでいたか?



視線方向と表情の組み合わせで対人関係がわかる

音声認識



NTT CS研 オープンハウス ×未来想論 2009

人間を知り、 情報の本質に迫る コミュニケーション科学 2009/6/4,5 オンライン開催

コンテンツを6月6日以降も公開します

京阪奈にご来場いただく 実地開催は中止となりました

トップページ

ごあいさつ プログラム

お問い合わせ

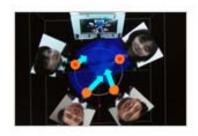
プログラム / 講演・テーマ展示一覧 /

テーマ展示

メディアとコミュニケーション B-1

会話の流れが一目瞭然!

コミュニケーションシーンを理解する音声映像技術



- |実時間会議モニタリング||一リアルタイム|&マルチモーダル対話シーン分析||--
- 激しく動いても見失わない顔追跡 ――複雑な動きに対して頑健な顔姿勢追跡――
- 会議も!談話も!実シーンの会話の文書化 ―会話音声認識・検索技術―!
- |会話の中から各話者の声を明瞭に聞き分けます||一残響に頑健な音声分離抽出技術―|
- | 爨いた収録音声を、くっきりとした音声に!|| ―映画/テレビ/CMの音声編集・調整用 残爨除去ソフト―

研究メンバー



大塚和弘



三上 弾



熊野史朗



大和淳司



荒木章子



石塚健太郎



藤本雅清



大庭隆伸

音声系

画像系

ご視聴ありがとうございました

NTT CS研 オープンハウス2009 オンラインサイト http://www.kecl.ntt.co.jp/openhouse/ 2009/theme/b1/index.html