

VNV研究会 2013/3/23

多人数インタラクション分析・理解のための マルチモーダル時系列データマイニング

東京工業大学 大学院総合理工学研究科

岡田 将吾

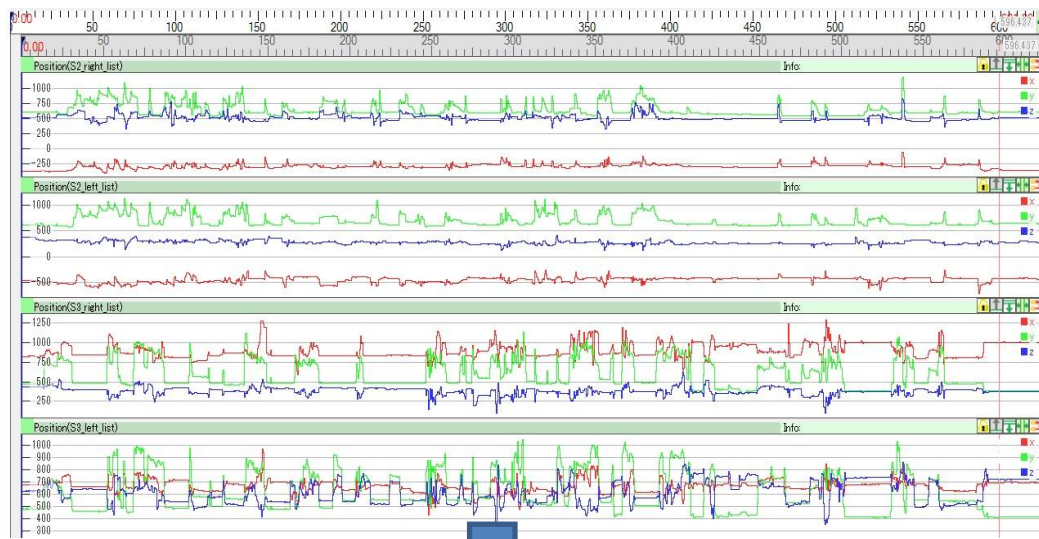
okada@ntt.dis.titech.ac.jp

- 研究背景:
- 研究目的:
- 関連研究
- 時系列データマイニングの概要
 - データ収集・アニメーション情景説明タスク
 - マルチモーダルデータの取得
 - ジェスチャデータの取得
 - 頭部動作・音声データの取得
 - 機械学習を用いた各モダリティデータのラベリング
 - マルチモーダルラベル系列からのパターン発見
- 時系列マイニングを利用した分析事例
 - データマイニングの結果を利用した分析結果
- まとめ

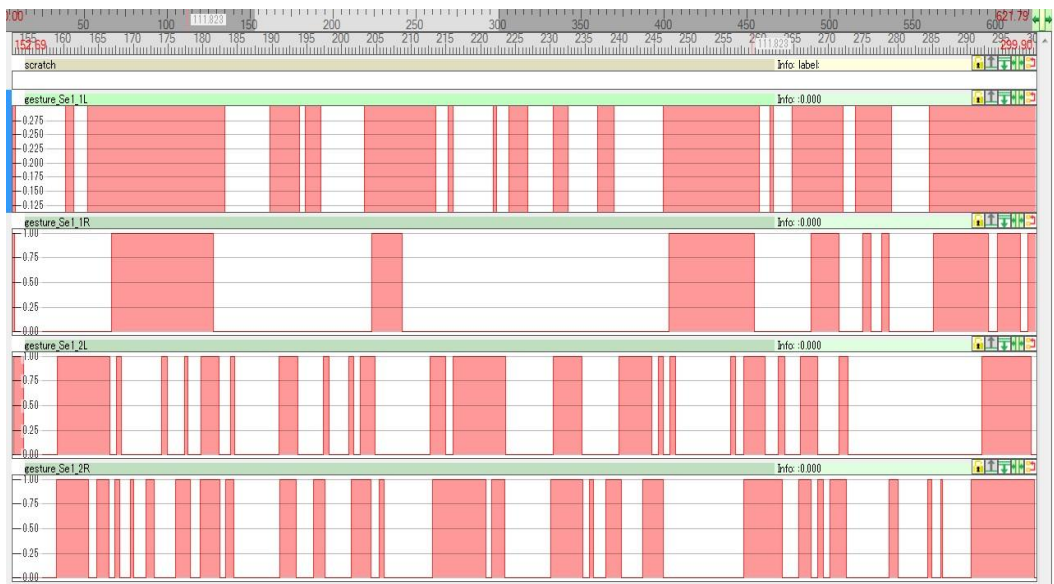
アニメーション説明タスク (①, ②が説明者、③が聞き手)



1. 説明者の手の動作データ取得

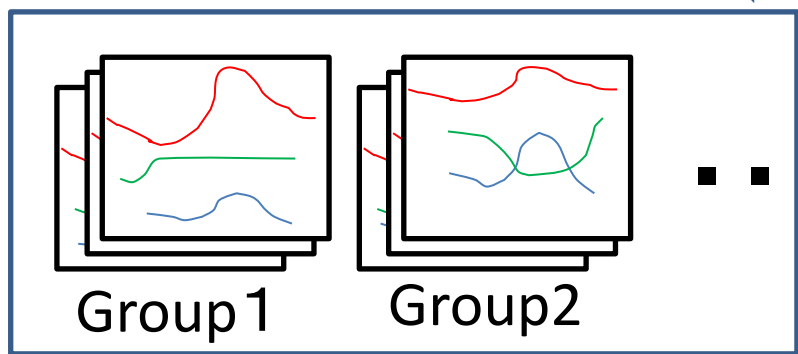


2. 説明者の手の動作区間抽出 (灰色の領域が動作区間)



デマインニング結果の提供

3. 頻出パターンが発見



本研究の背景・動機

データ中心科学的手法の開発が重要

- メリット1: 大量データから客観的な指標を獲得可能
- メリット2: 分析者の負担を軽減可能
- メリット3: ビデオ映像分析から気付けなかった規則性の発見

データ中心科学的手法を実現するためには……

① 対面会話状況で交わされる非言語表現と音声言語表現とを統合的に収録出来る環境

⇒ IMADEルーム (インタラクション分析を目的としたセンサルーム)

② 収録した大規模なデータを効率よく自動(半自動)で分析するための環境

⇒ 非言語データを処理可能なマイニング手法の開発(本研究の目的)

本研究の背景・動機

データ中心科学的手法の開発が重要

- メリット1: 大量データから客観的な指標を獲得可能
- メリット2: 分析者の負担を軽減可能
- メリット3: ビデオ映像分析から気付けなかった規則性の発見

データ中心科学的手法を実現するためには……

① 対面会話状況で交わされる非言語表現と音声言語表現とを統合的に収録出来る環境

⇒ IMADEルーム (インタラクション分析を目的としたセンサルーム)

② 収録した大規模なデータを効率よく自動(半自動)で分析するための環境

⇒ 非言語データを処理可能なマイニング手法の開発(本研究の目的)

本研究の背景・動機

コミュニケーション中に会話参加者から表出される 非言語パターン

韻律 (発話有無)

顔表情

手のジェスチャ

姿勢

視線

頭部ジェスチャ

⋮

本研究の背景・動機

会話におけるジェスチャの重要性

韻律 (発話有無)

顔表情

手のジェスチャ

姿勢

視線

頭部ジェスチャ

ジェスチャ研究の進展

(代表例1) ケンドンによるジェスチャ単位

(代表例2) マクニールによるキャッチメント
理論・成長点理論



ジェスチャは発話・談話と密接な関係を持つ



会話中のジェスチャの機能を分析・モデリングするための情報技術の構築が重要

本研究ではハンドジェスチャのマイニングツール構築に焦点をあてる。

関連研究

非言語パターン・会話構造のマイニング

AMI, CHIL プロジェクト

- ・ 非言語行動 (例 ターンテイキングの認識) の自動認識
- ・ AMI: 会話環境はミーティング形式に特化
- ・ CHIL: オフィスや教室におけるインタラクションを収録

IDIAP 研究所

- ・ 視線, 頭部ジェスチャ, 発話などから会話の発話権をよくにぎっている人を特定
- ・ 表出されている非言語パターンとお互いの印象・個性・タスク達成率の間の相関分析

関連研究

非言語パターン・会話構造のマイニング

大塚 大和ら(NTTグループ)

- 複数人物による対面会話における, 会話の構造推定を実施
- 会話現象を確率的な事象とみなし, 観測された非言語行動と会話構造との関係を表す確率モデル(階層ベイズモデル)を構築
- CVでの顔トラッキングの高速化・高精度化や, 顔の表情を認識する方法を提案

中田・角 (IMADEグループ 京大)

- 発話・相槌・指さし・視線・頷きを手動でラベリング
- N gram 統計のイベント列の抽出に基づくデータマイニングを実行
- ポスター発表会話と展示物に対する自由会話における
会話構造を抽出・比較

本研究で扱う、「ハンドジェスチャに関するマイニング」は行われていない

関連研究

ジェスチャの機械学習・認識

工学系のジェスチャに関する多くの研究では……

- 認識対象のジェスチャの**種類や数を限定**、
- **大量の**教師ラベル付きデータによってモデルを構成

(Mitra, S.; Acharya, T., “Gesture Recognition: A Survey,” IEEE SMC Part C, 2007)

VACEプロジェクト……

- 対面ミーティングにおける会話構造の特定を目的とする
- マクニールが加わり、自身のキャッチメントや成長点といった概念を取得されたデータで確かめた
- 多人数の会話場面でもジェスチャの理論が適用できる可能性を示唆している

・多人数コミュニケーション中でのジェスチャの役割分析・分類は未解決
⇒センサデータからのデータマイニングのアプローチで解決したい(解決に近づきたい)

本研究の動機

問題点・研究の動機

・多人数コミュニケーション中での非言語インタラクション・特にジェスチャの役割分析・分類は未解決

⇒センサデータからのデータマイニングのアプローチで解決に近づきたい

本研究の仮説

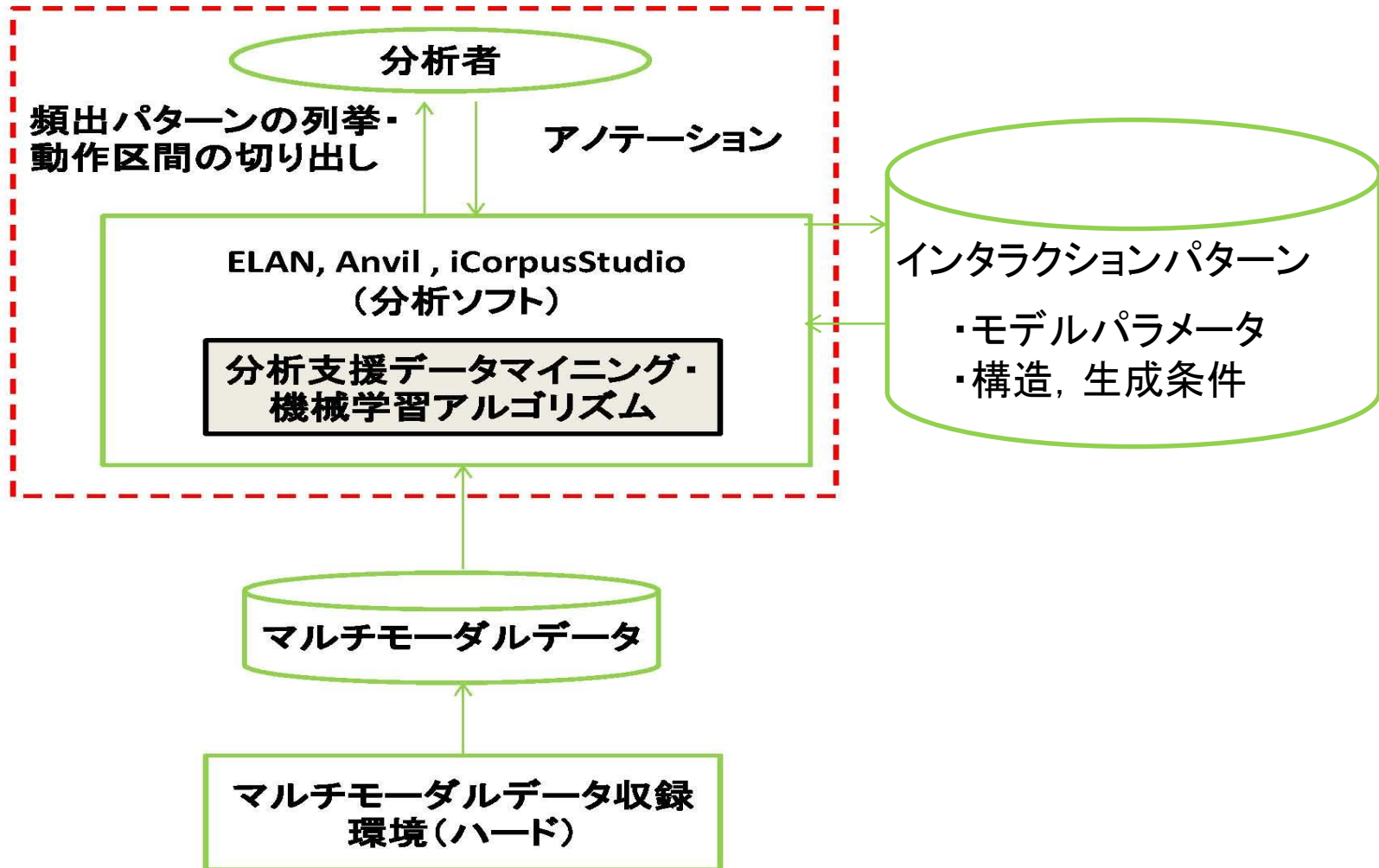
インタラクション時の相手の様子や、会話の状態をジェスチャと同時に観測することが、会話中に産出されるジェスチャを理解することに近づく

本研究の目的

1. ジェスチャ分析者が獲得した知見・仮説を基に、大規模会話データの分析を行うためのデータマイニング手法の開発
2. 頭部動作の認識から、首振り動作・顔向け動作の自動ラベリング手法を開発
3. マルチモーダルパターンから会話状態を推定し、各会話状態で用いられたジェスチャパターンを分析

本研究の動機

本研究の枠組み
(分析援助のために機械学習が存在)



- 研究背景:
- 研究目的:
- 関連研究
- **時系列データマイニングの概要**
 - データ収集・アニメーション情景説明タスク
 - マルチモーダルデータの取得
 - ジェスチャデータの取得
 - 頭部動作・音声データの取得
 - 機械学習を用いた各モダリティデータのラベリング
 - マルチモーダルラベル系列からのパターン発見
- 時系列マイニングを利用した分析事例
 - データマイニングの結果を利用した分析結果
- まとめ

- 研究背景:
- 研究目的:
- 関連研究
- **時系列データマイニングの概要**
 - **データ収集・アニメーション情景説明タスク**
 - マルチモーダルデータの取得
 - ジェスチャデータの取得
 - 頭部動作・音声データの取得
 - 機械学習を用いた各モダリティデータのラベリング
 - マルチモーダルラベル系列からのパターン発見
- 時系列マイニングを利用した分析事例
 - データマイニングの結果を利用した分析結果
- まとめ

対象とするタスク: アニメーション再生課題 「Canary Row」を説明するタスク

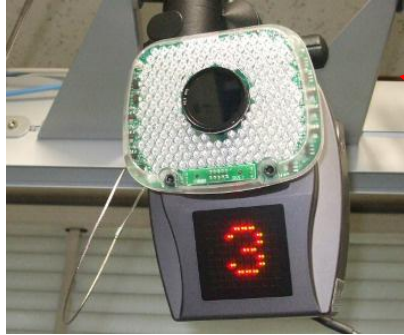


利点:

- 1 たくさんのジェスチャが表出される
- 2 言語での説明内容がある程度固定されている
- 3 過去のジェスチャ研究と比較が可能となる
- 4 説明者を二人とすることで、協調的説明・主導的な説明シーンなど様々な説明スタイルを観測出来る

- 研究背景:
- 研究目的:
- 関連研究
- **時系列データマイニングの概要**
 - データ収集・アニメーション情景説明タスク
 - **マルチモーダルデータの取得**
 - **ジェスチャデータの取得**
 - **頭部動作・音声データの取得**
 - 機械学習を用いた各モダリティデータのラベリング
 - マルチモーダルラベル系列からのパターン発見
- **時系列マイニングを利用した分析事例**
 - データマイニングの結果を利用した分析結果
- **まとめ**

非言語情報をセンシングするデバイス



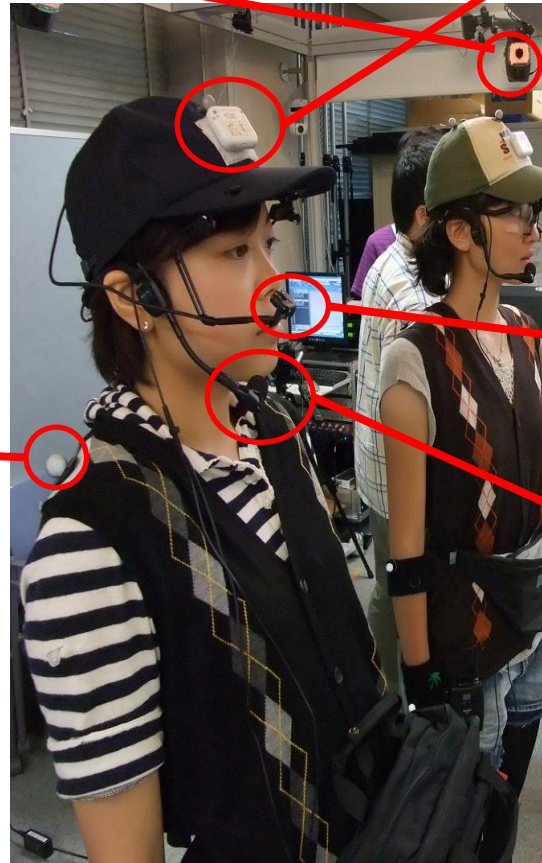
ラプターカメラ



マーカ

光学式モーションキャプチャシステム (NAC社製)

腕部・頭部の動作取得



加速度センサ
(ATR Promotion
社製)

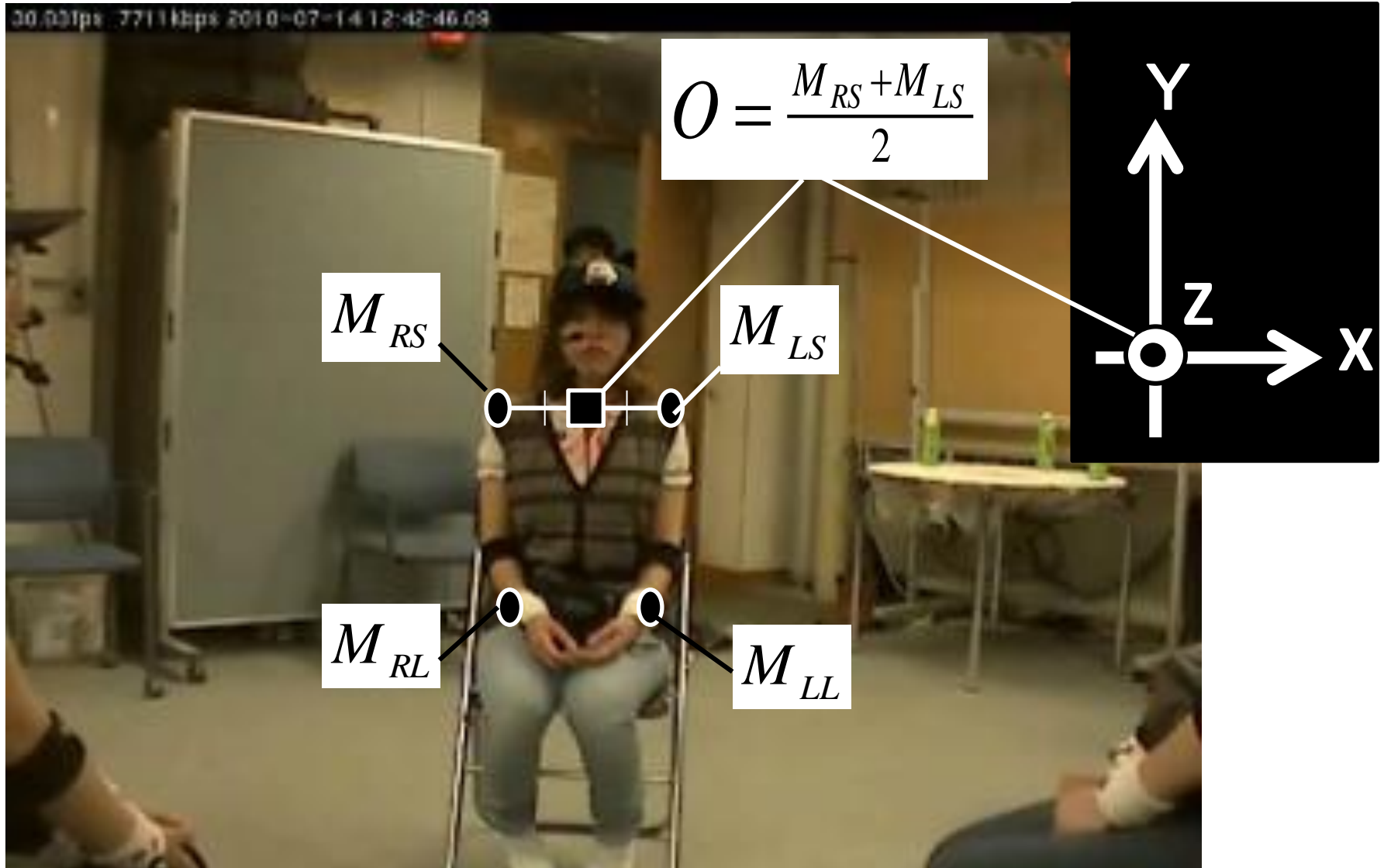
頭部の動作取得

アイトラッカ
(NAC社製)

指向性マイク
(Shure 社製)

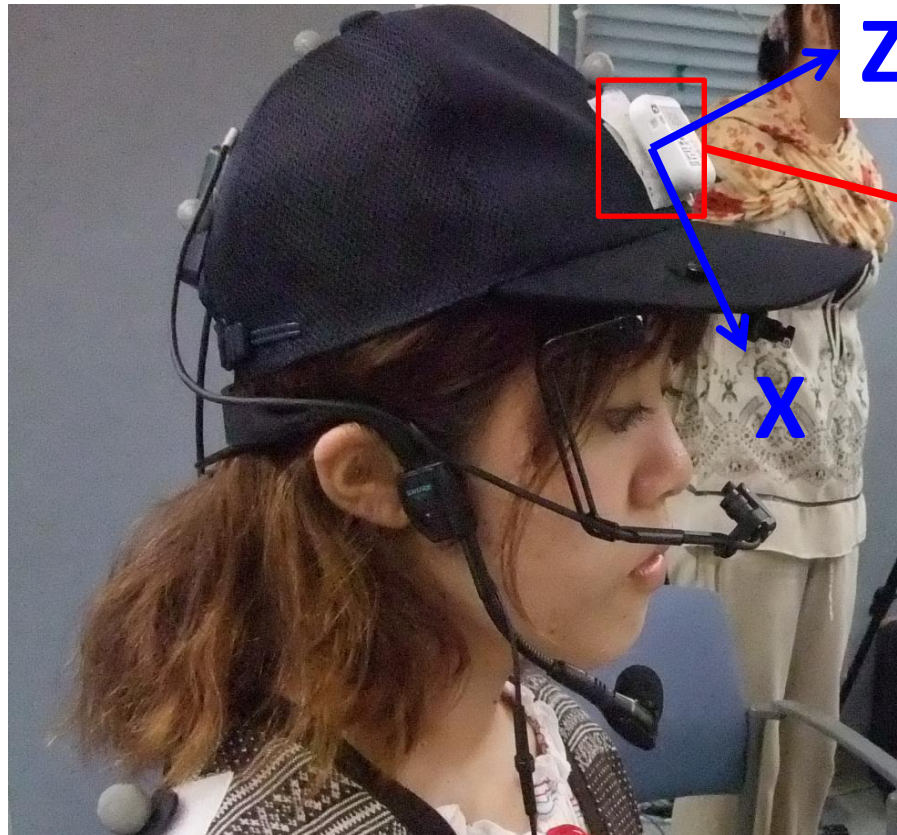
発話有無の取得

ジェスチャパターン取得のための座標変換



M_* , O は3次元の位置座標

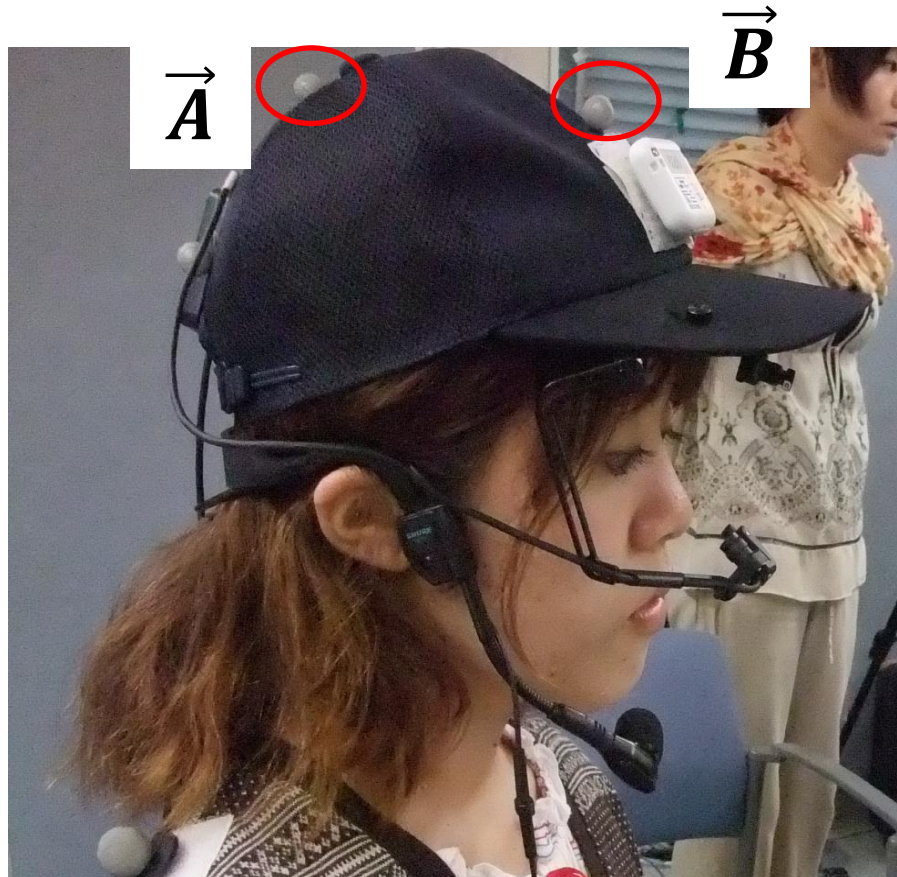
頭部動作取得のための情報取得
: 頭部の縦方向動作 (うなづきを含む)



加速度データ取得

縦方向の頭部動作検出のために
加速度のX, Z成分を利用

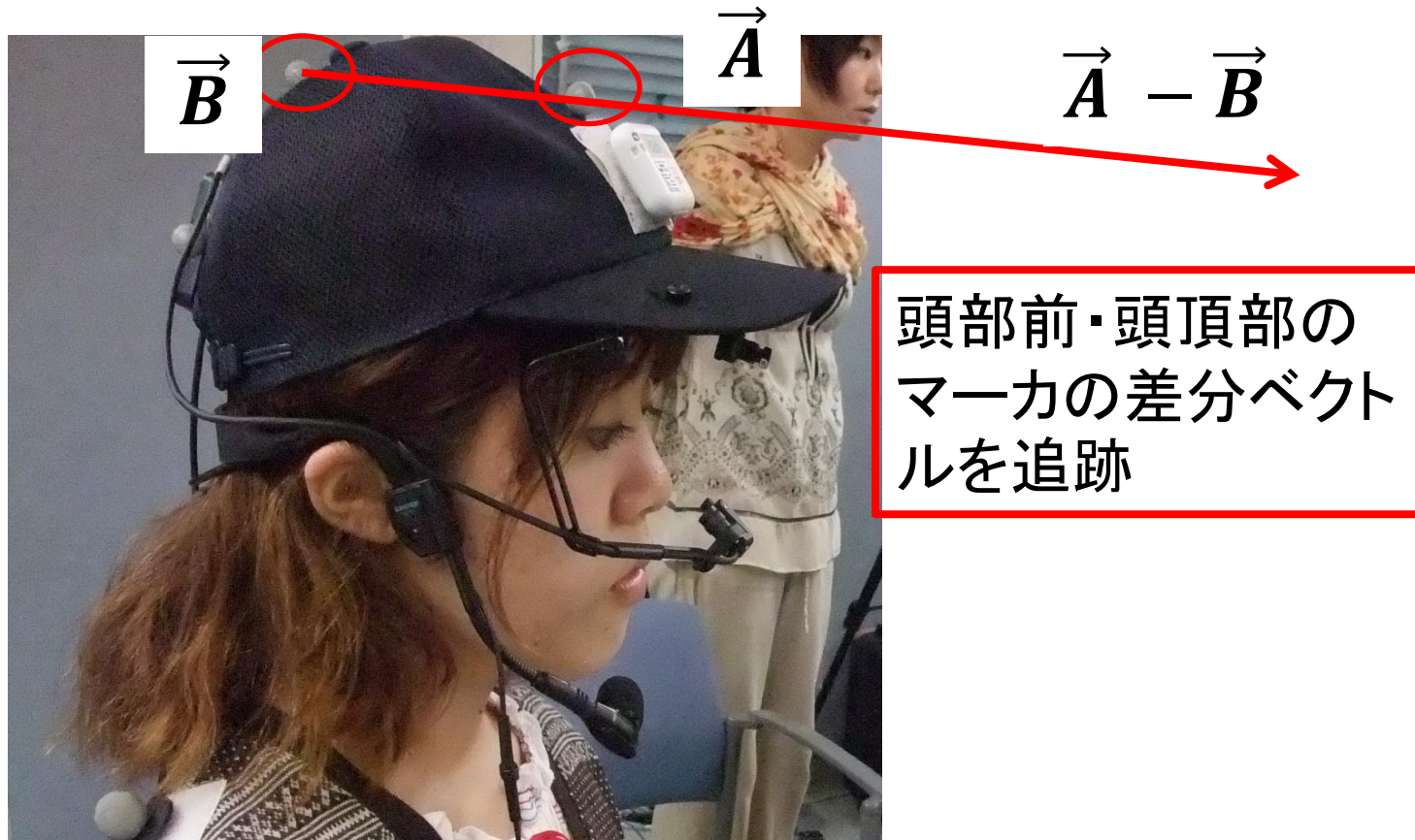
頭部動作取得のための情報取得 : 頭部の横方向の動作 (顔向け動作)



座標データ取得

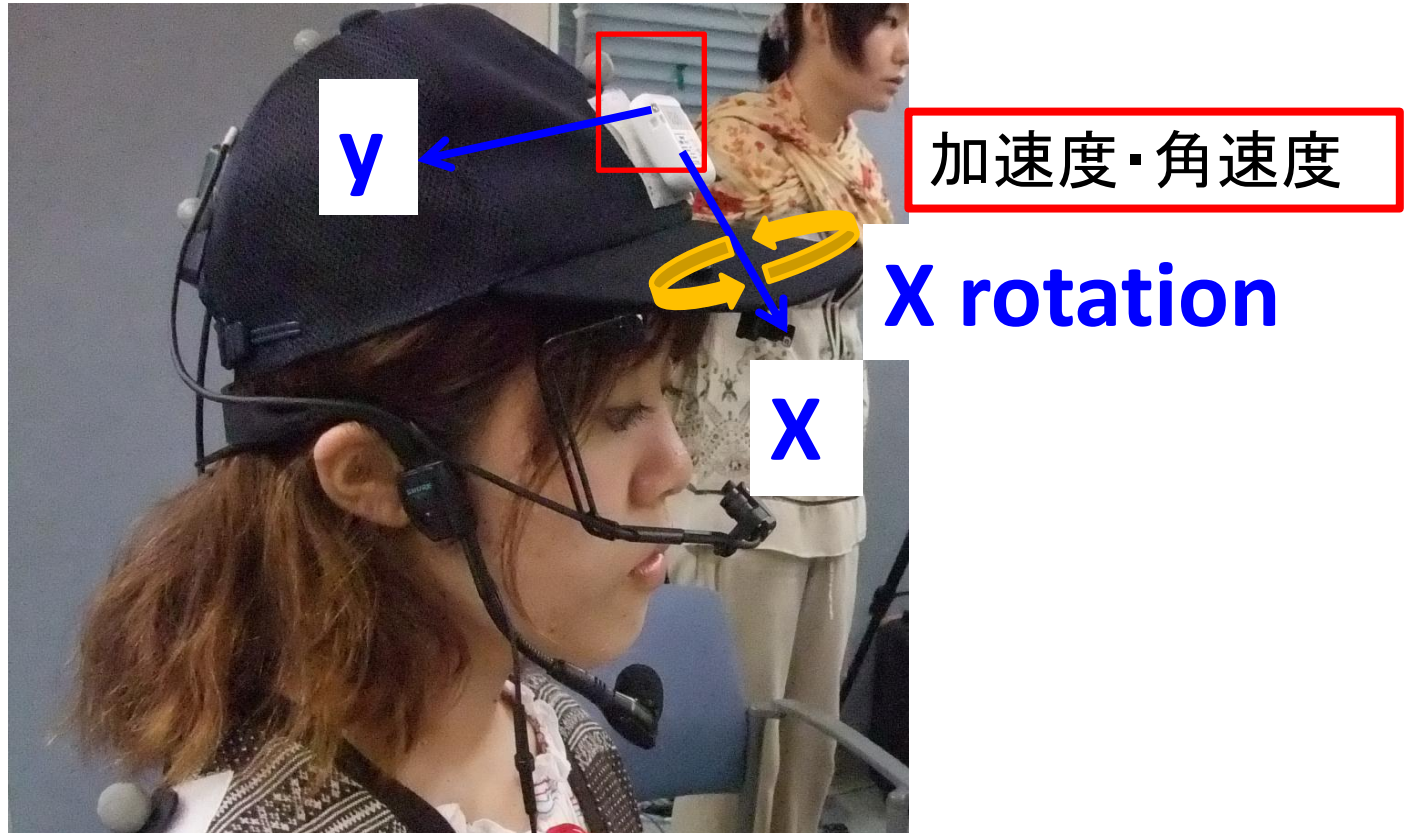
横方向の頭部動作検出のために
モーションセンサの頭部マーカの位置座標

頭部動作取得のための情報取得 : 頭部の横方向の動作 (顔向け動作)



横方向の頭部動作検出のために
モーションセンサの頭部マーカの位置座標を利用

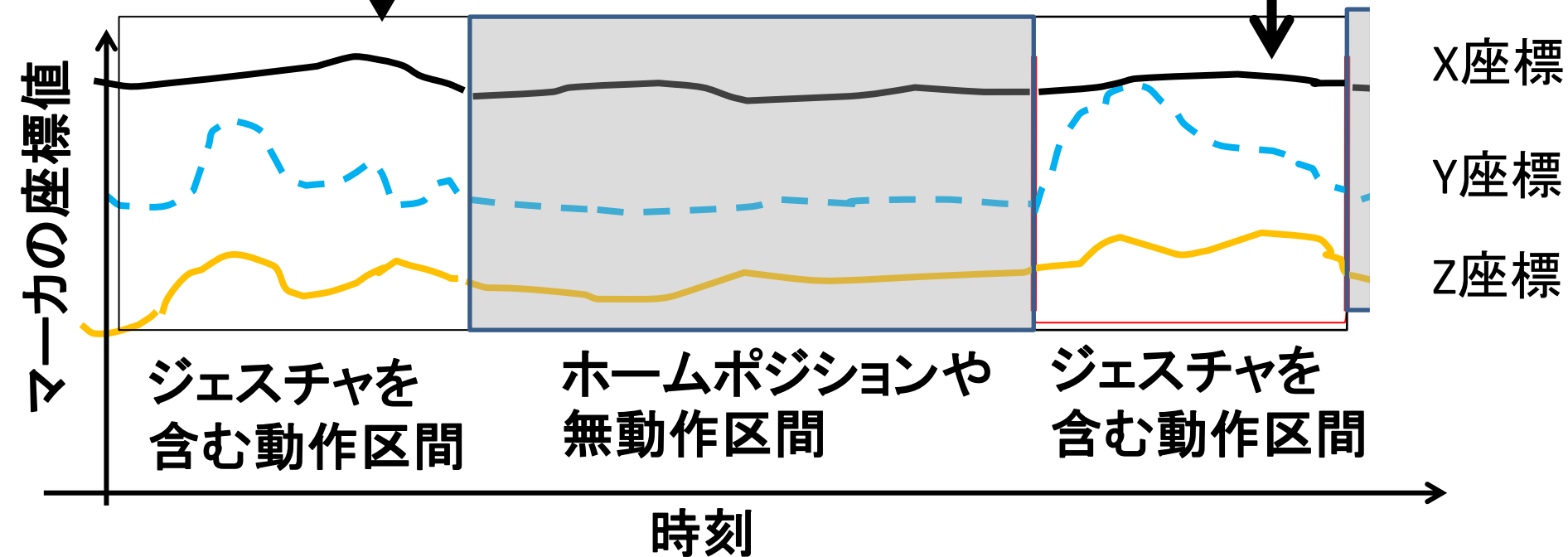
頭部動作取得のための情報取得 : 頭部の横方向の動作 (顔向け動作)



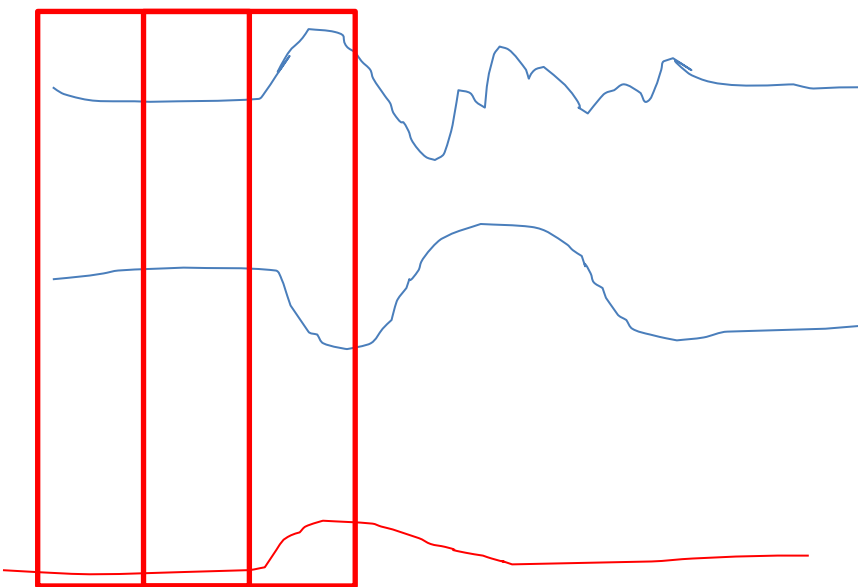
頭部マーカの観測値が欠損している場合,
加速度センサのy成分, x軸周りの角速度を利用可能

- 研究背景:
- 研究目的:
- 関連研究
- **時系列データマイニングの概要**
 - データ収集・アニメーション情景説明タスク
 - マルチモーダルデータの取得
 - ジェスチャデータの取得
 - 頭部動作・音声データの取得
 - **機械学習を用いた各モダリティデータのラベリング**
 - マルチモーダルラベル系列からのパターン発見
- 時系列マイニングを利用した分析事例
 - データマイニングの結果を利用した分析結果
- まとめ

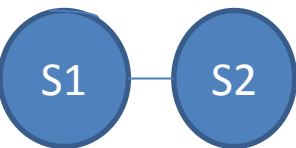
腕部動作検出のためのセグメンテーション



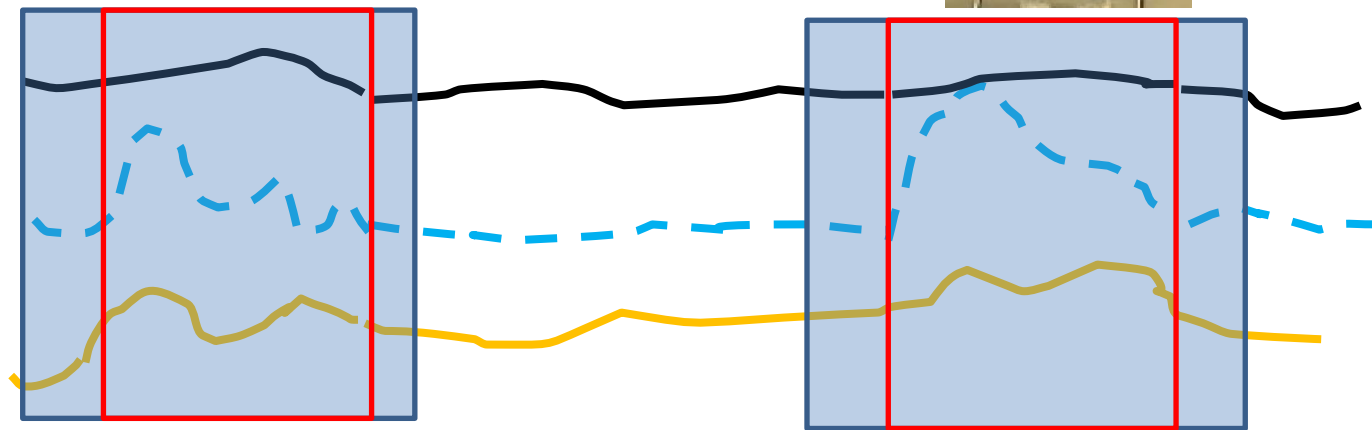
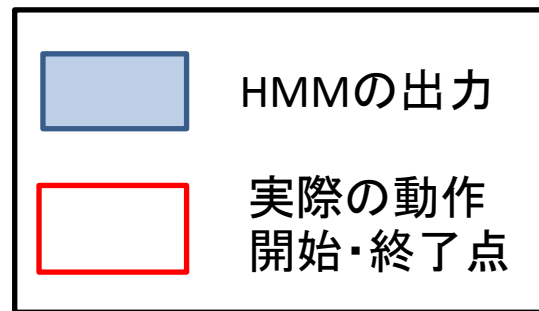
腕部動作セグメンテーションのための手法： HMM (Hidden Markov Model)



1. 無動作領域・動作領域から HMM1, HMM2を学習
2. 一定の窓枠に対し2つの HMMからの尤度を算出
3. 2つのHMMの尤度 $L1, L2$ の比をスコアと定義
4. 閾値を上回れば動作と判定
 $L2(\text{動作})/L1(\text{無動作}) > th$

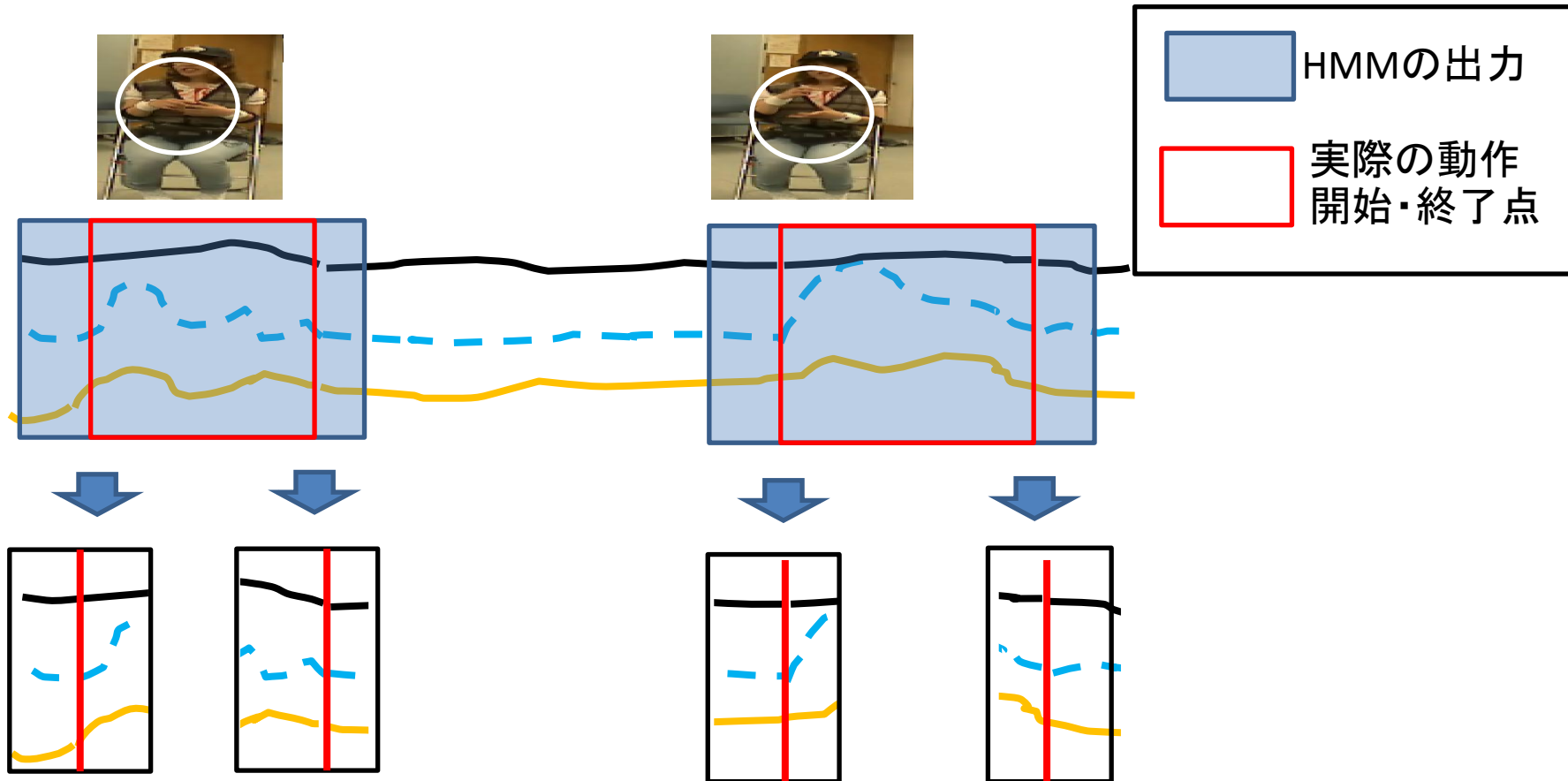


腕部動作セグメンテーションのための手法： HMMの問題点と解決法



- ・HMMは分節点を正確に切り出せない
⇒HMMの出力した境界領域の前後で最も急激に変化する点を検出し、分節点をアップデートすることで対処

セグメンテーションのための手法： SSTによる変化点検知手法の導入



- ・変化点検知手法SST (Singular Spectrum Transform [Ide 2005])を利用
- ・HMMが出力した動作領域の始点・終点付近で一番変化が大きい点を検出し、この点を動作の分節点とする。

腕部動作セグメンテーション精度の比較結果

- S6を訓練データとして, S1のデータをテスト
- センサデータが欠損しているところは除く
- 正解セグメント境界と±10フレーム(30f/s)であれば正解

	適合率	再現率	F値
Y座標の差分に基づく分節	0.72	0.83	0.80
SWAB	0.88	1.00	0.94
HMM	0.97	1.00	0.98
HMM+SST	0.99	0.99	0.99

頭部動作検出のためのセグメンテーション 対象とする動作



横方向の動き

- ・顔向け動作はアドレスを示す

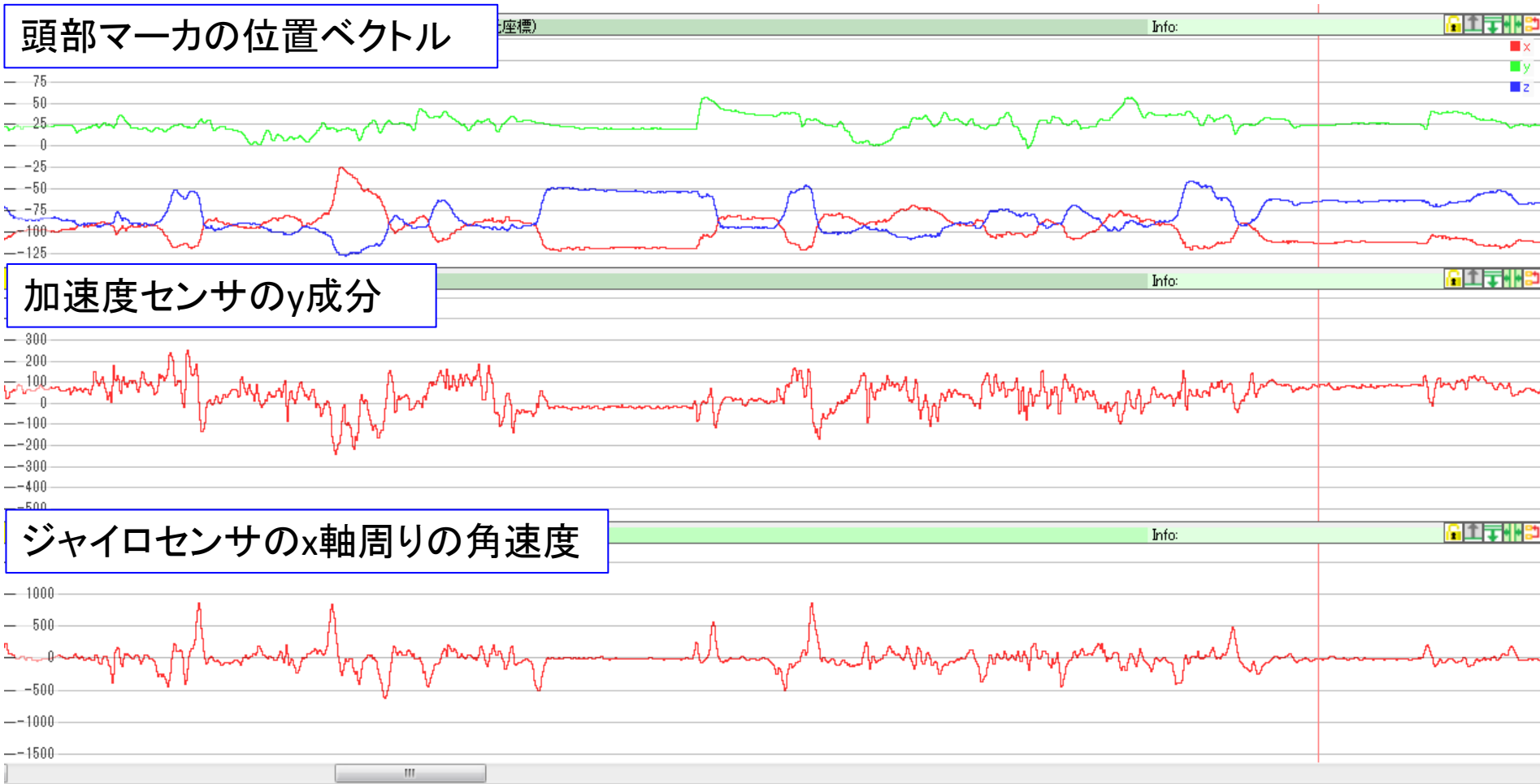


縦方向の動き

- ・うなづき (同意:理解)
- ・発話の調子を整える

横方向の頭部動作のラベリング:

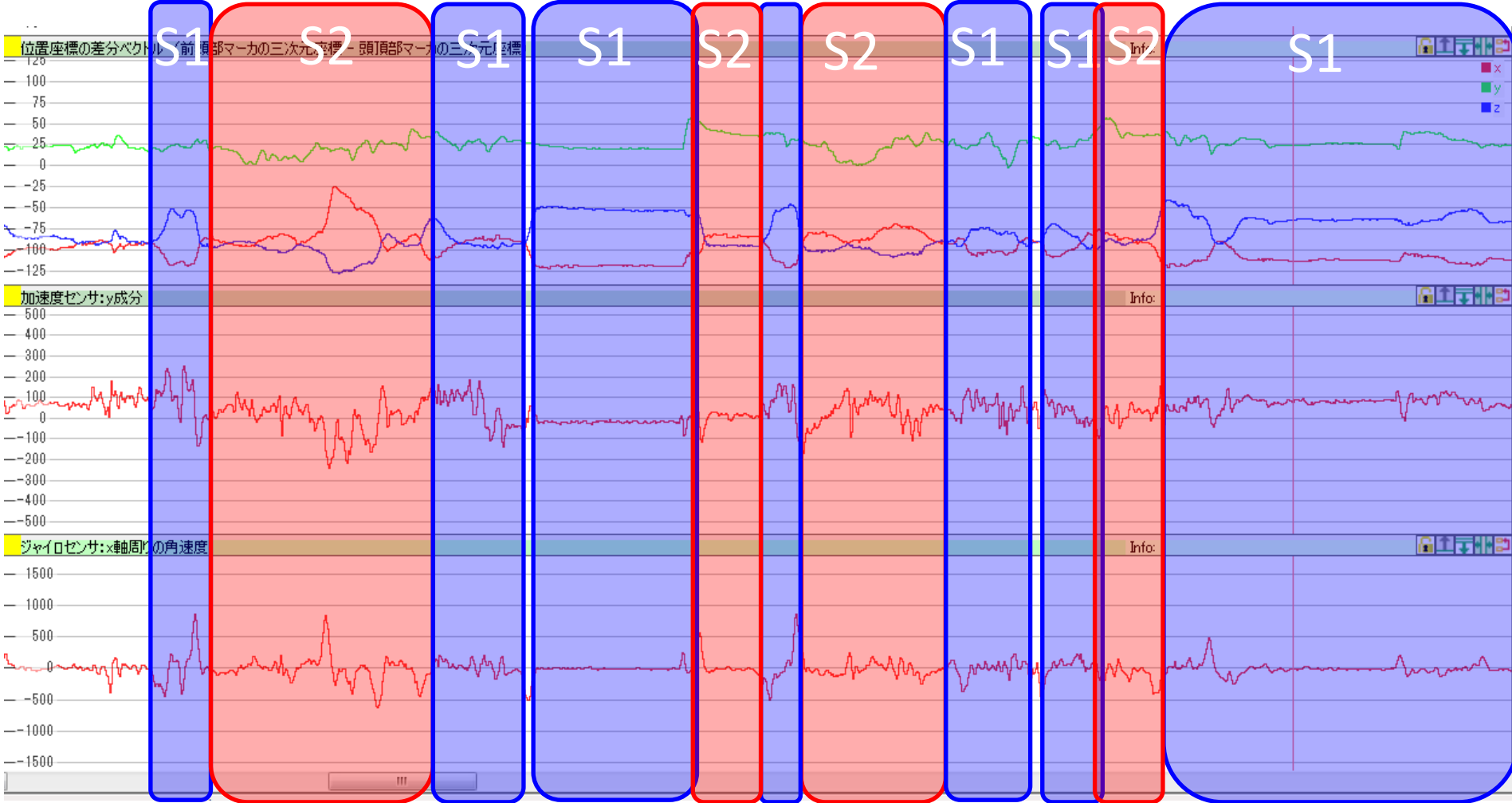
Speaker 3の頭部マーカ-加速度センサ-ジャイロセンサの値



- タスクの特性上, 参照物は存在しない
- 顔向けの方は当人以外の2人のどちらかに近似

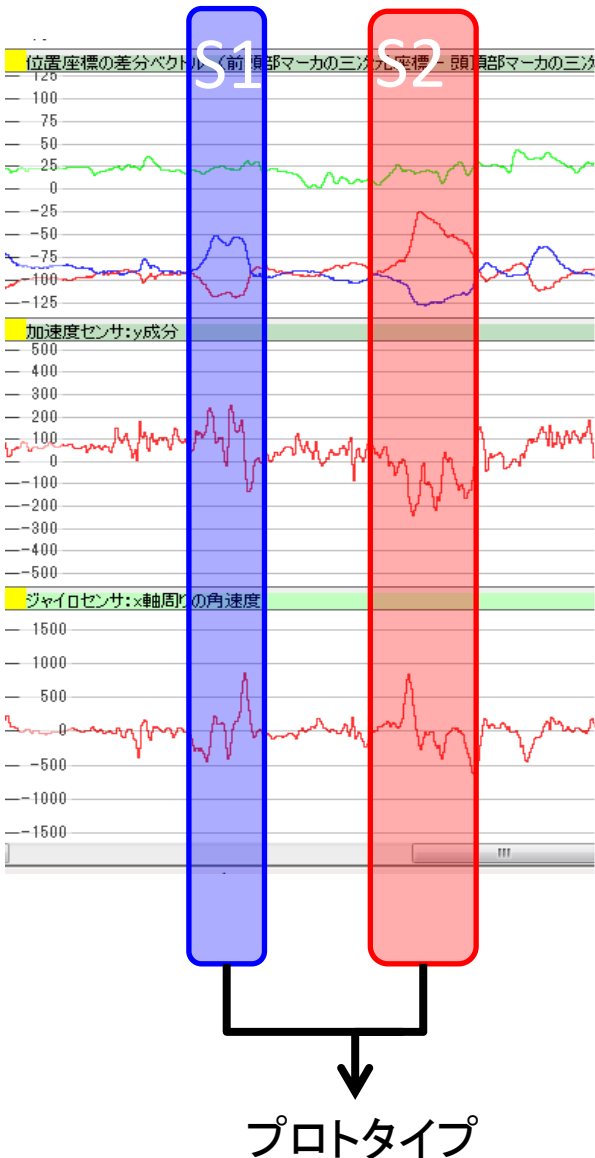
横方向の頭部動作のラベリング:

Speaker 3の頭部マーカ-・加速度センサ・ジャイロセンサの値



- ・タスクの特性上, 参照物は存在しない
- ・顔向けの方は当人以外の2人のどちらかに近似

横方向の頭部動作のラベリング手法: KNN



識別器作成フェーズ

1. 各ラベルごとにプロトタイプ(代表)データを収集
2. 差分・平均・分散などの特徴量を抽出, ベクトル化

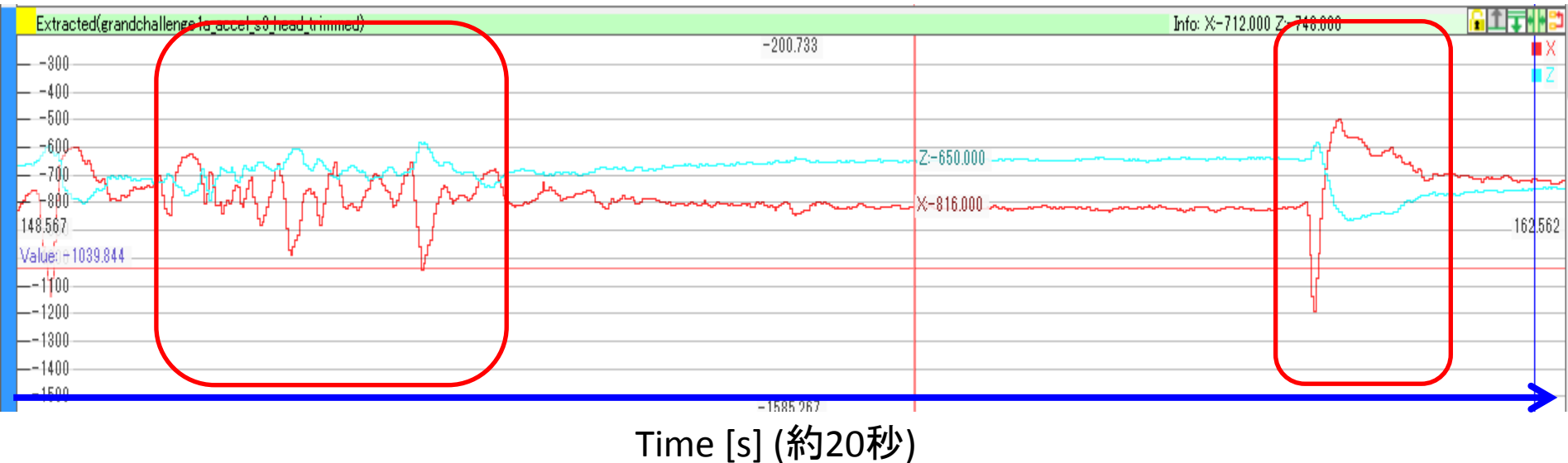
ラベリングフェーズ

1. 信号データを適当な窓幅で切り取り, 入力データセットを作成
2. プロトタイプ1NN(最近傍識別)によって入力データをラベリング

縦方向の頭部動作のラベリング : SVM

Speaker 3の加速度センサ

加速度センサのx,z 成分 (mv/g)

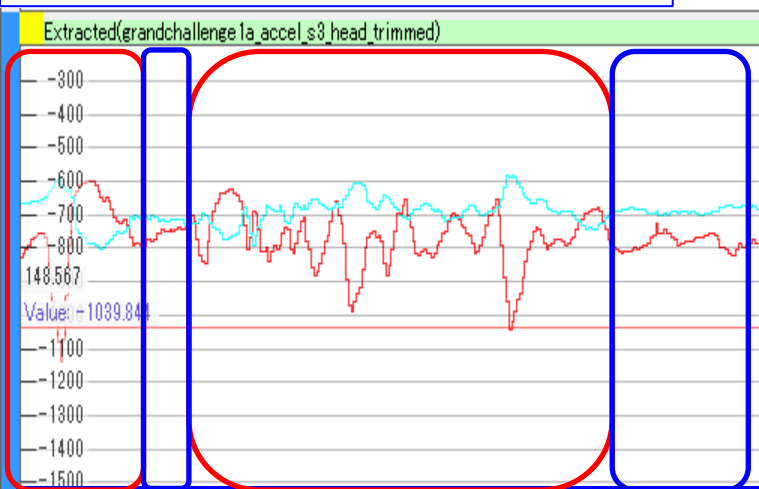


- ・頭部の縦振り動作の検出
⇒聞き手の場合ほうなづきである可能性が高い

縦方向の頭部動作のラベリング手法:SVM

Speaker 3の加速度センサ

加速度センサのx,z 成分 (mv/g)



識別器作成フェーズ

1. 各ラベルごとにプロトタイプ(代表)データを収集
2. Wavelet変換を行い, 高・低周波数成分に分解
3. 差分・平均・分散などの特徴量を抽出
4. SVMで頭部動作パターンの識別器作成

ラベリングフェーズ

1. 信号データを適当な窓幅で切り取り, 入力データセットを作成
2. SVMで入力データをラベリング

頭部動作有

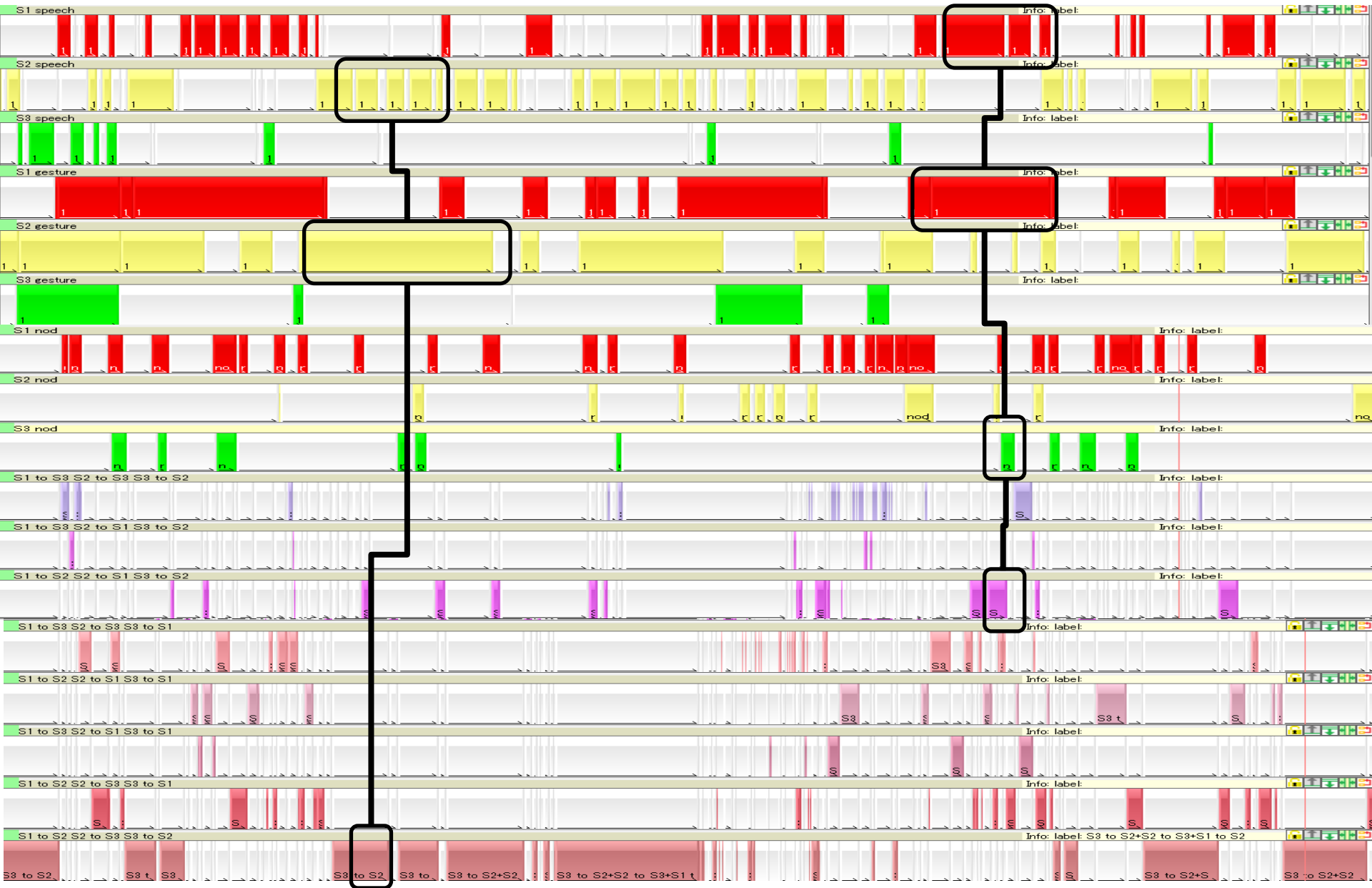
頭部動作無

- 研究背景:
- 研究目的:
- 関連研究
- **時系列データマイニングの概要**
 - データ収集・アニメーション情景説明タスク
 - マルチモーダルデータの取得
 - ジェスチャデータの取得
 - 頭部動作・音声データの取得
 - 機械学習を用いた各モダリティデータのラベリング
 - **マルチモーダルラベル系列からのパターン発見**
- 時系列マイニングを利用した分析事例
 - データマイニングの結果を利用した分析結果
- まとめ

自動ラベリングにより獲得されたパターン

モダリティ	ラベリング手法	精度 (適合率)	獲得が期待できるインタラクション 行為・インタラクション行為理解の ための役割
頭部動作 (横方向)	KNN	1セッションでテスト 75~80%程度 (改善必要あり)	<ul style="list-style-type: none">・話者・聞き手の顔向け・アドレス・インタラクションイベント (相互注視)
頭部動作 (縦方向)	Linear SVM	1セッションでテスト 70%程度 (改善必要あり)	<ul style="list-style-type: none">・聞き手のうなづき・話し手の発話調整
腕部動作区 間	HMM+SST	90%以上	<ul style="list-style-type: none">・ジェスチャ
発話有無	GMM+ ゼロ点交差法	90%以上	<ul style="list-style-type: none">・話者の特定

インタラクションイベントの発見：概要図

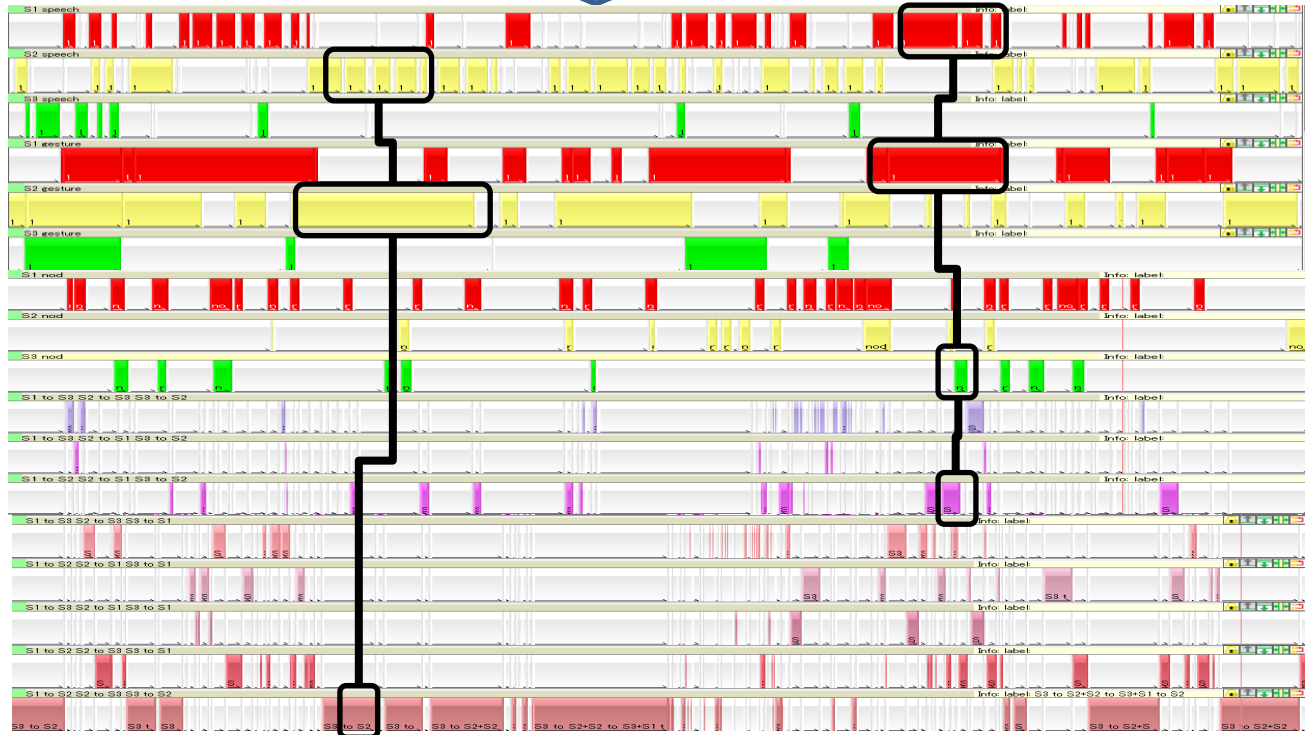
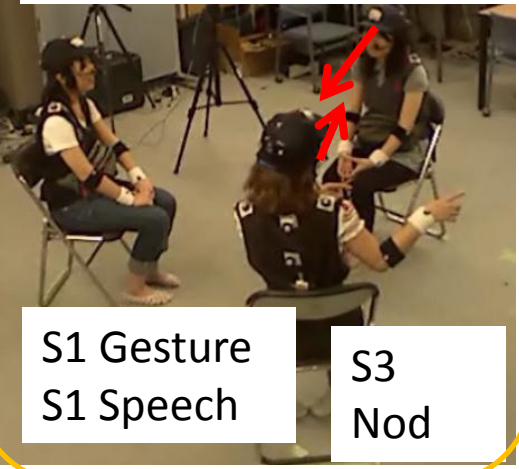


インタラクションイベントの発見：概要図

各モダリティのデータのラベリング

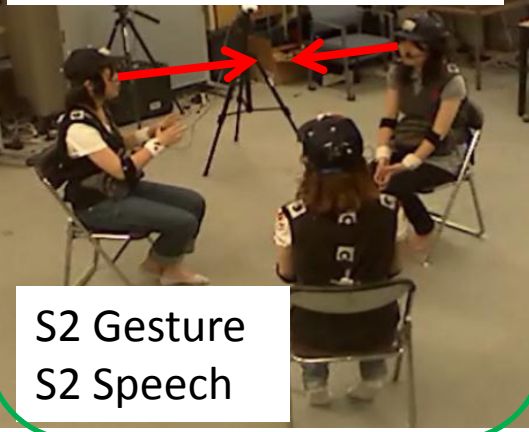


S1 (S3) faces to S3 (S1)



頻出イベントパターン

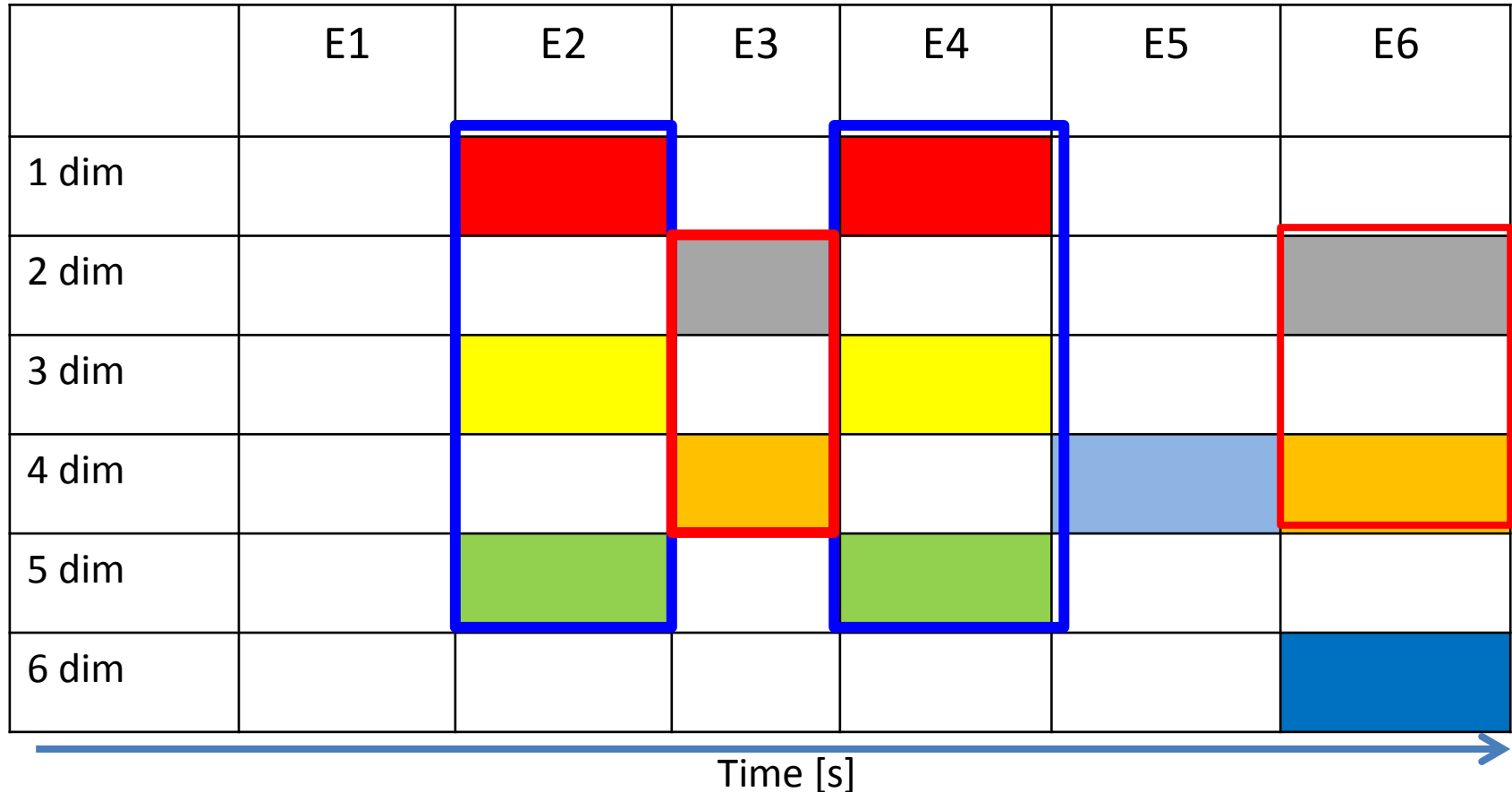
S2 (S3) faces to S3 (S2)




インタラクションイベントを多次元共起パターンとみなす


マルチモーダルイベント発見における課題1

従来法 (例えば, Minnen 2007)



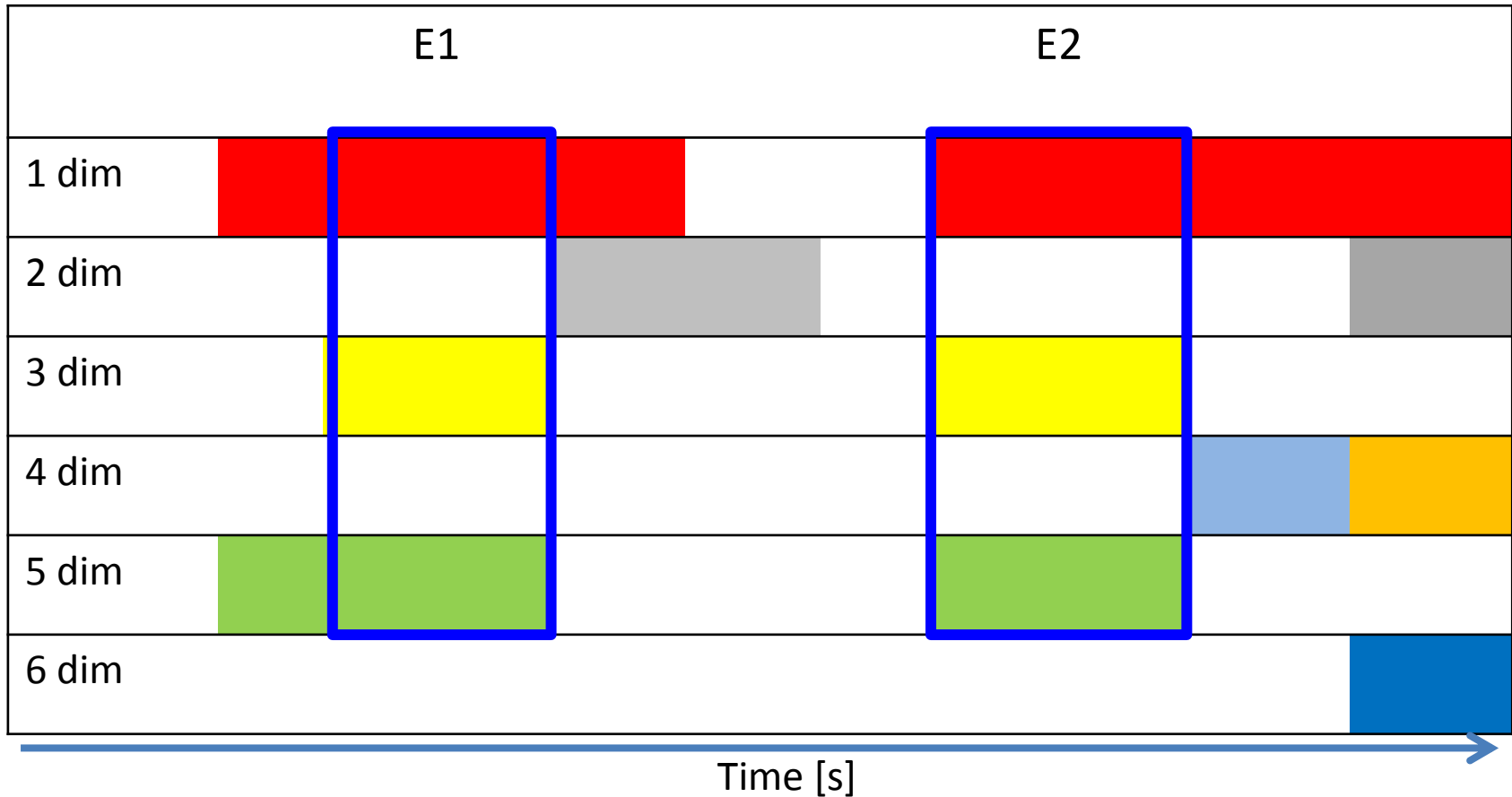
• 従来手法では,  のようなすべての次元にまたがる同期パターンを抽出可能



⇔  のような部分的な構造一致を抽出することに焦点を当てていなかった.

※E6の6 dimで  のイベントが起きているため, E6とE3は別の状態として扱われる

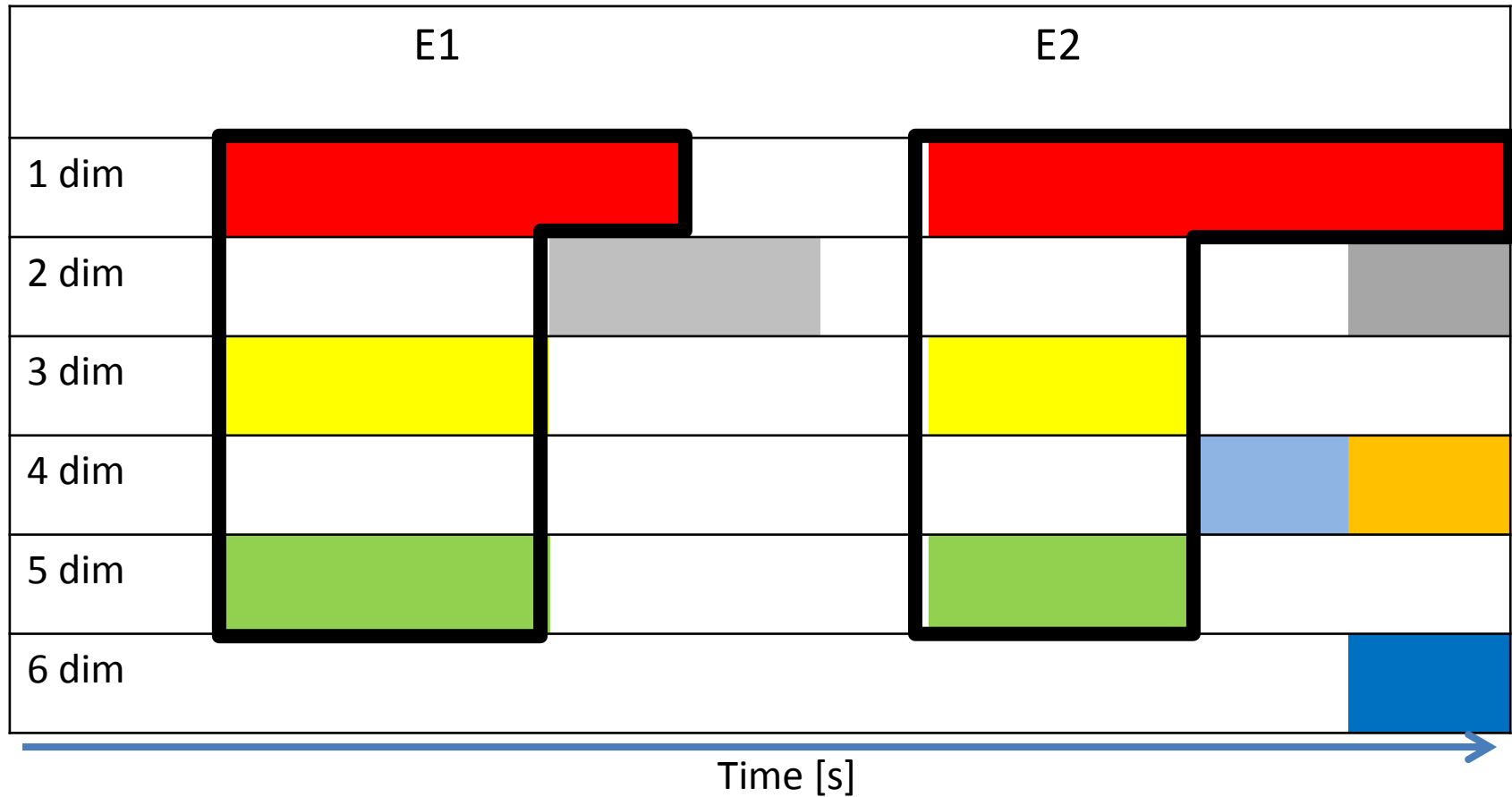
マルチモーダルイベント発見における課題2

従来法 (例えば, Minnen 2007)



- 従来手法では、のようなすべての次元にまたがる同期パターンを抽出可能
- ⇔ のような部分的な構造一致を抽出することに焦点を当てていなかった。

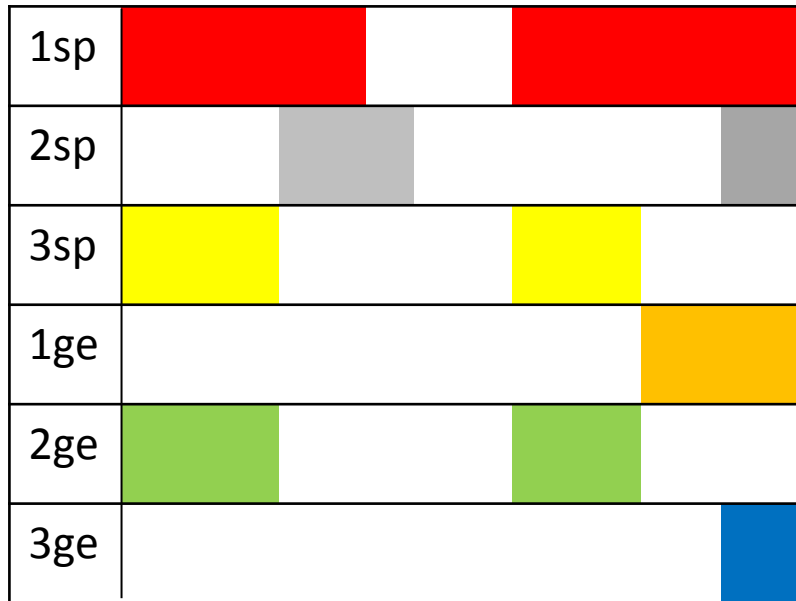
マルチモーダルイベント発見手法 (Alireza 2009)



1. パターンのタイミング差を許容して共起パターンを抽出
2. 部分空間パターンも抽出可能

マルチモーダルイベント発見手法 (Alireza 2009)

マルチモーダルラベルデータ



	1sp	2sp	3sp	1ge	2ge	3ge
1sp	*	1	1	0.5	1	0.5
2sp	1	*	0	0.5	0	0
3sp	1	0	*	0	1	0
1ge	1	1	0	*	0	1
2ge	1	0	1	0	*	0
3ge	1	1	0	1	0	*

Spは発話, geはジェスチャ

1. パターンのタイミング差を許容して共起パターンを抽出
2. 部分空間パターンも抽出可能

マルチモーダルイベント発見手法 (Alireza 2009)

アルゴリズム

1. 各パターンの共起関係から共起行列を計算
2. パターンの多い順にソート
3. パターン p_i と共起度が th 以上のパターンを統合
4. 統合されたパターン v_j を i 番目のパターン集合 S_i に追加
5. v_j を所属していた集合 S_j から除外
6. v_j と p_i の共起パターンをさらに新しいパターンとして登録
7. 共起度行列を更新
8. Step3の条件を満たすパターンがなくなるまで step2から7を繰り返す

(Alireza 2009)の手法のデメリット

アルゴリズム

1. 各パターンの共起関係から共起行列を計算
2. **パターンの多い順にソート**
3. パターン p_i と共起度が th 以上のパターンを統合
4. 統合されたパターン v_j を i 番目のパターン集合 S_i に追加
5. v_j を所属していた集合 S_j から除外
6. v_j と p_i の共起パターンをさらに新しいパターンとして登録
7. 共起度行列を更新
8. Step3の条件を満たすパターンがなくなるまで step2から7を繰り返す

数が一番多いパターンが少数のパターンを統合する

⇒数が多いパターンが必要なパターンとは限らない

(例: 説明者の発話と聞き手のうなづきの共起パターンなど)

本研究における(Alireza 2009)の手法の拡張

CMPDアルゴリズム

1. 各パターンの共起関係から共起行列を計算
2. **統合順序に制約を加える**
3. パターン p_i と共起度が th 以上のパターンを統合
4. 統合されたパターン v_j を i 番目のパターン集合 S_i に追加
5. v_j は, 所属していた集合 S_j から除外
6. v_j と p_i の共起パターンをさらに新しいパターンとして登録
7. 共起度行列を更新
8. Step3の条件を満たすパターンがなくなるまで step2から7を繰り返す

分析対象となるジェスチャラベルの優先度を上げて, パターン数が少なくても共起度が大きいパターンの優先度を次に高くする. ⇒ Constrained Multi modal Pattern Discovery と呼称

- 研究背景:
- 研究目的:
- 関連研究
- 時系列データマイニングの概要
 - データ収集・アニメーション情景説明タスク
 - マルチモーダルデータの取得
 - ジェスチャデータの取得
 - 頭部動作・音声データの取得
 - 機械学習を用いた各モダリティデータのラベリング
 - マルチモーダルラベル系列からのパターン発見
- 時系列マイニングを利用した分析事例
 - データマイニングの結果を利用した分析結果
- まとめ

データマイニング手法を援用した量的分析

- 三人組（1人が説明・1人が聞き手）のアニメーション課題を8セッション収録
- 8セッションで収録されたセンサ時系列データセットから以下のデータマイニングを実施
 - （1）頭部・腕部動作有無，発話有無を検出
 - （2）頻出インタラクションパターン発見
- データマイニング結果を分析

センサデータの自動ラベリング結果

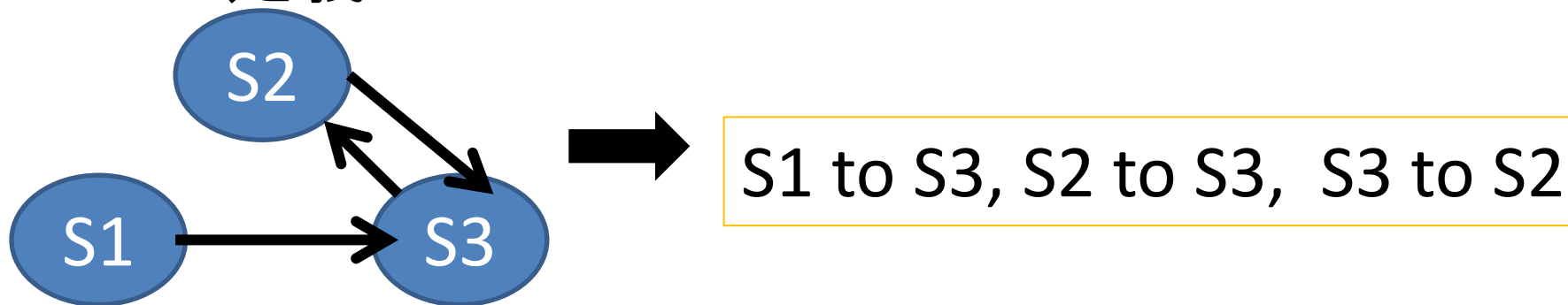
- 8セッションで収録されたセンサ時系列データセットに対し、頭部動作・腕部動作・発話ラベルを自動付与した
- 700ms以下の短い発話パターンは除外した

	腕部 動作	頭部 縦動作	頭部 横動作	発話
パターン 総数	499	2584	10270	4426

マルチモーダルイベント発見の手順 1

頭部横ラベルについて

1. 話者X looks 話者Yとしてラベルを付与
2. 全参加者のラベルをマージし, 会話中の顔向け状態として定義



各ラベルの取り扱い

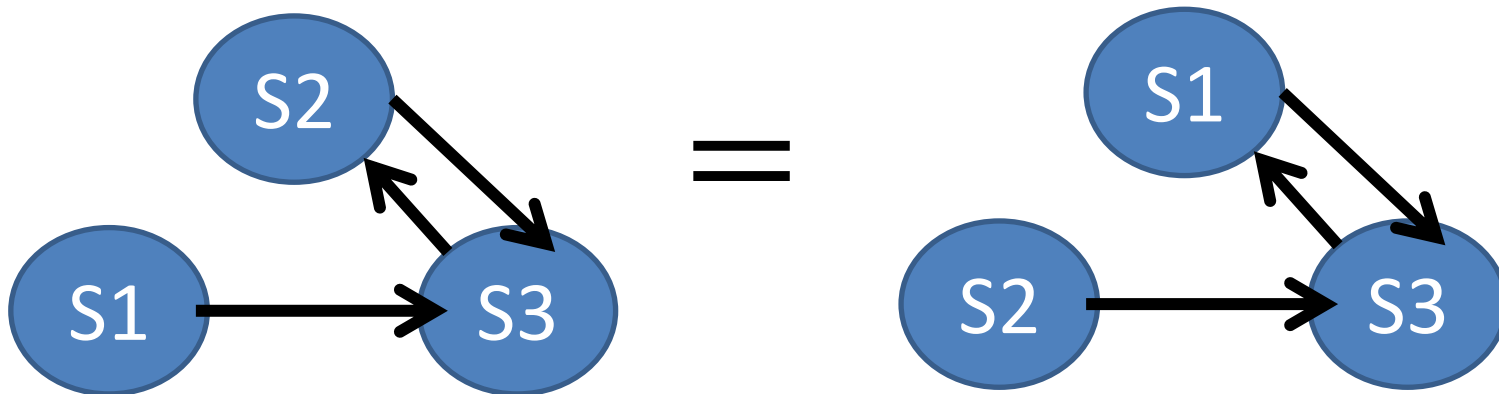
- 全モダリティのラベルはon またはoffとする.
- 3話者の発話・腕部動作・頭部縦動作 (計9次元), 顔向け状態 (計6次元) を合わせて15次元準備

マルチモーダルイベント発見の手順2

会話役割の考慮

- アニメーション説明課題では、説明者2人と聞き手1人の役割は明確に分かれている

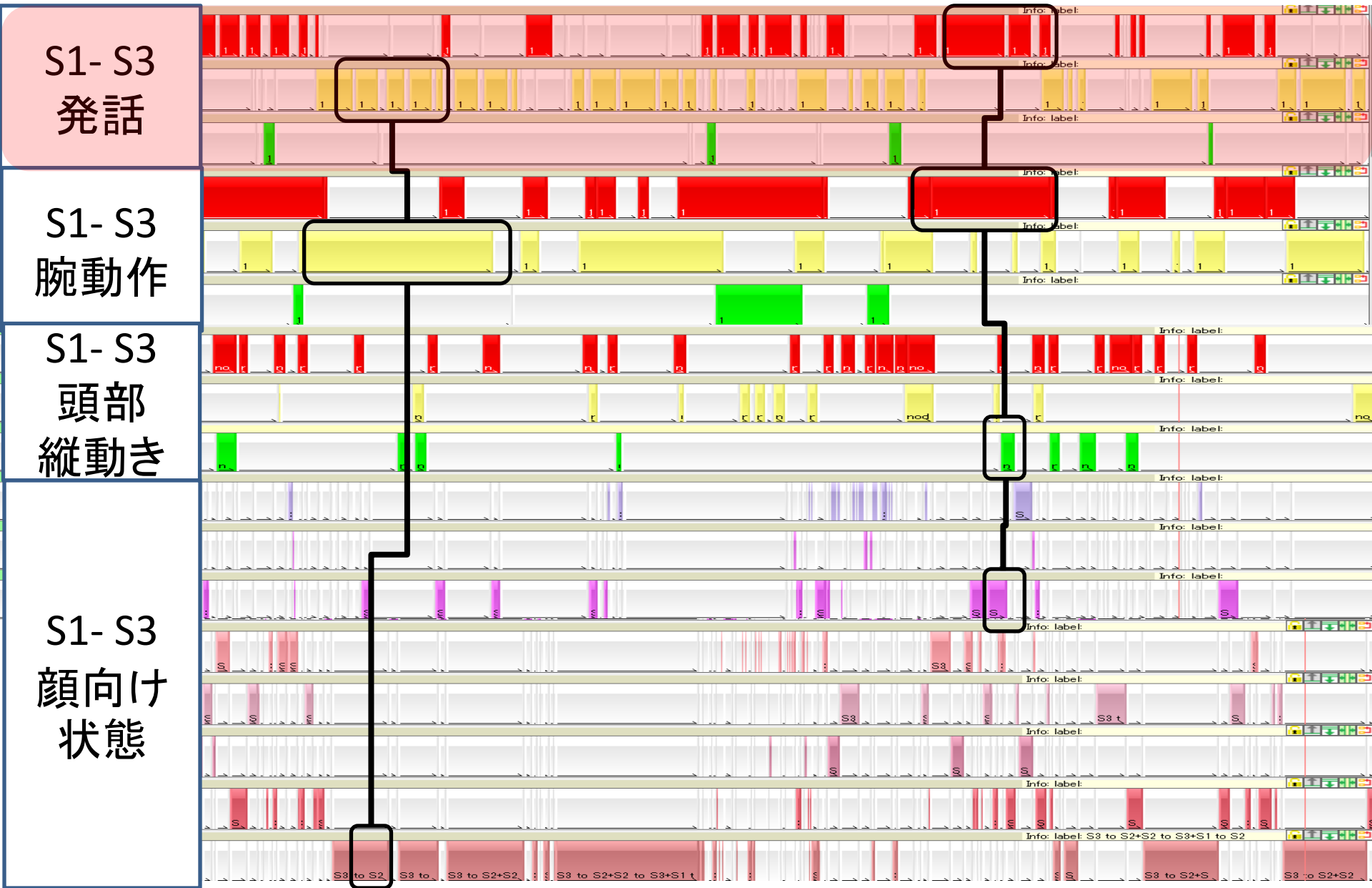
⇒ 説明者1, 2から見た, 各状態を統一するために両方を起点とした時系列をCMPDへの入力とする.



CMPDの閾値

総数の内, 30%の個数が共起していれば, マルチモーダルパターンと認定

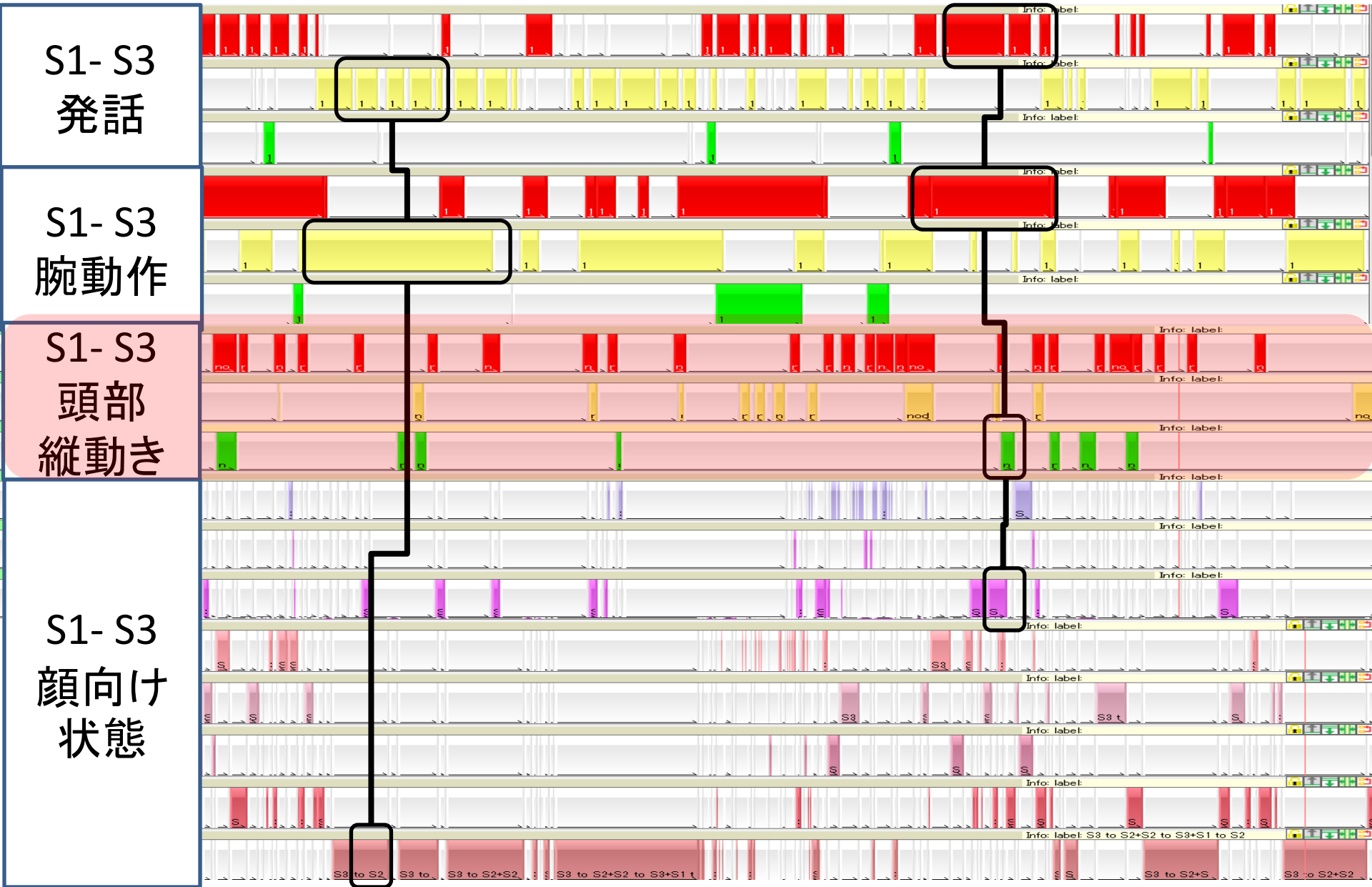
マルチモーダルイベント列の例



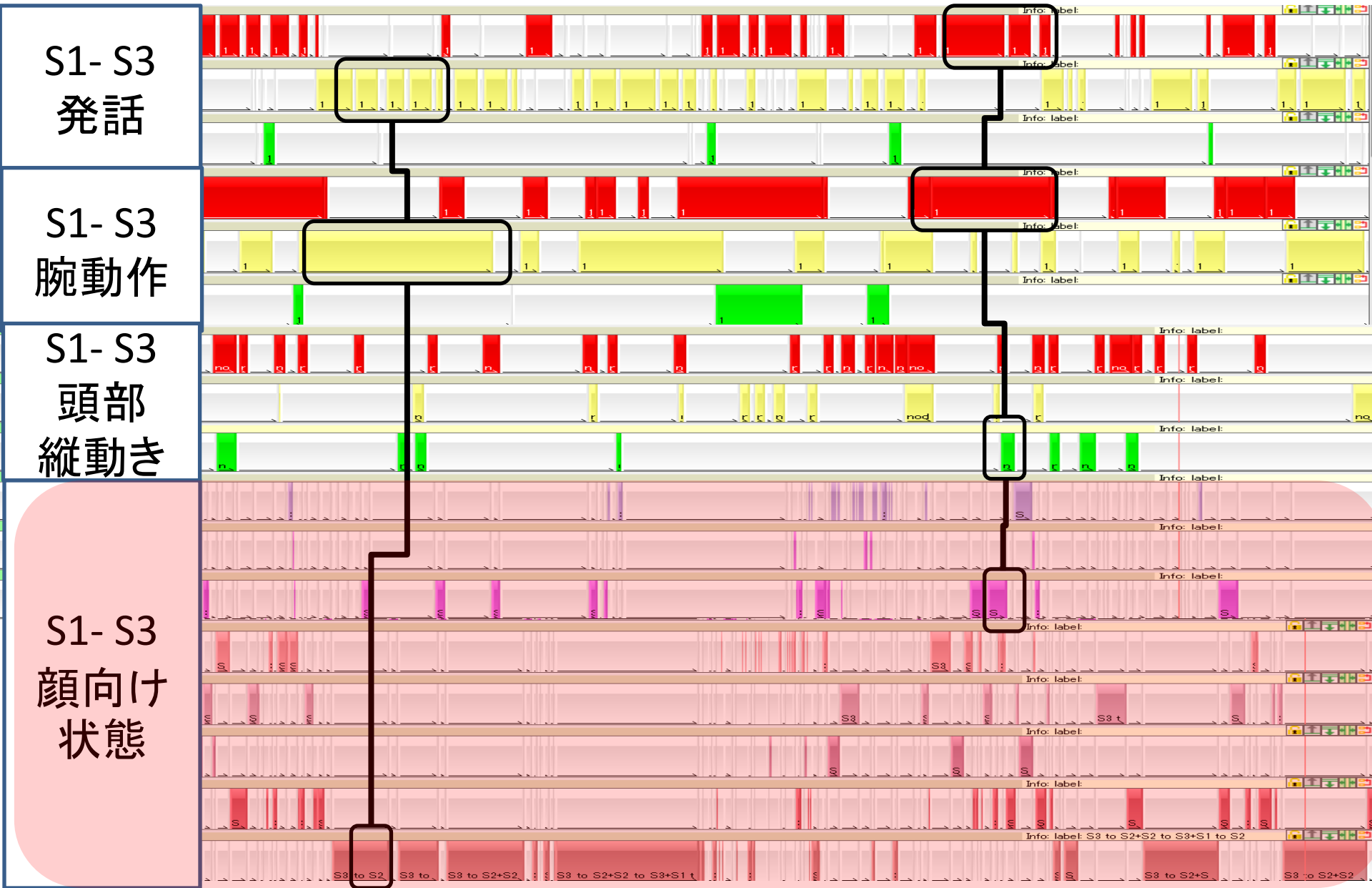
マルチモーダルイベント列の例



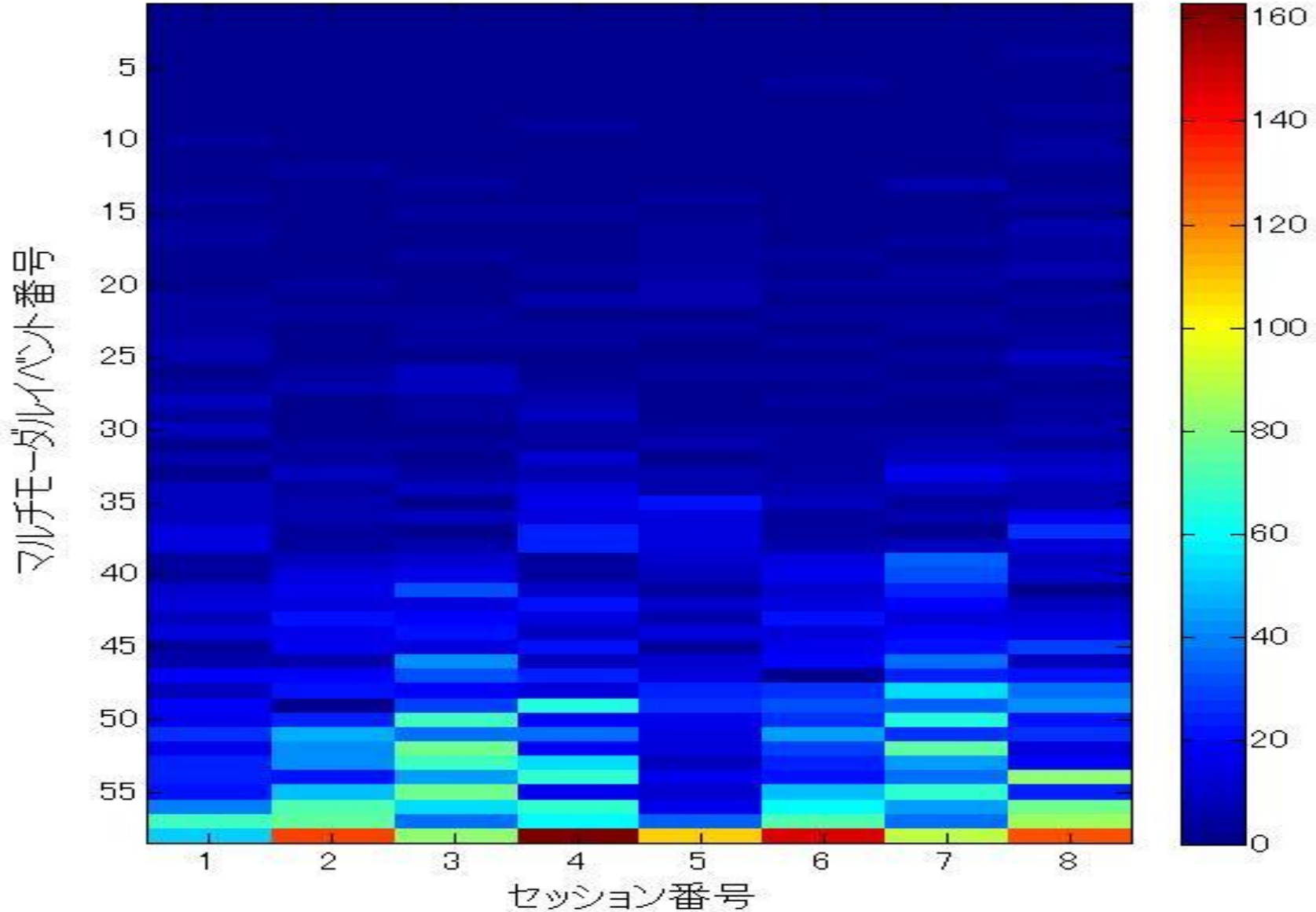
インタラクションイベントの発見：概要図



マルチモーダルイベント列の例

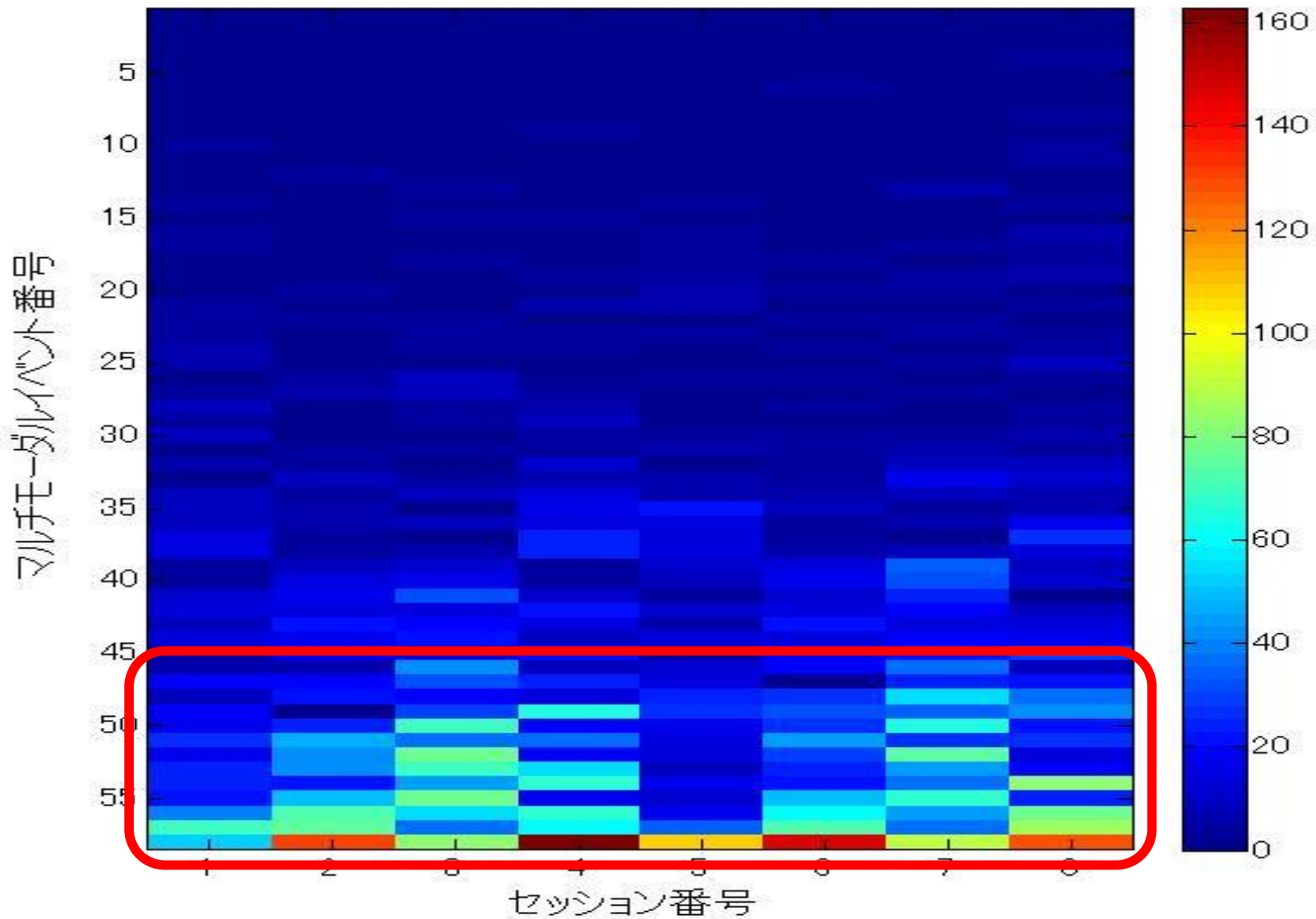


マルチモーダルイベント発見の結果



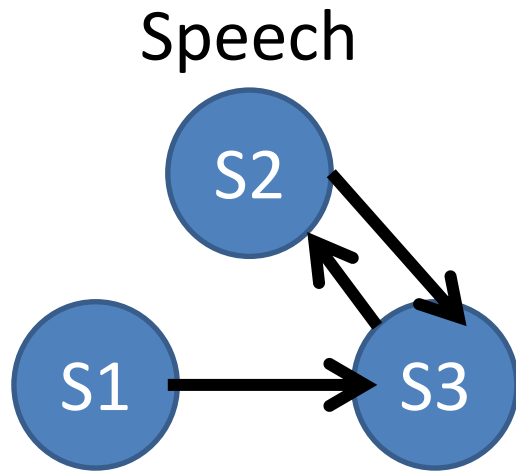
58個のパターンが抽出された。(図は総数が多い順にソートしてある)

マルチモーダルイベント発見の結果

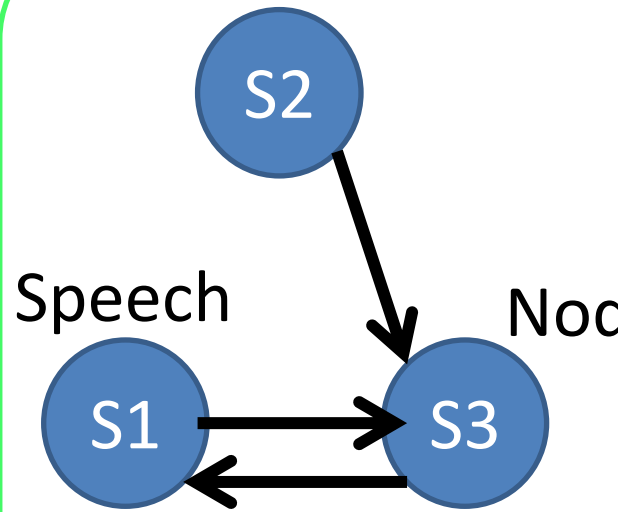


58個のパターンが抽出された。(図は総数が多い順にソートしてある)

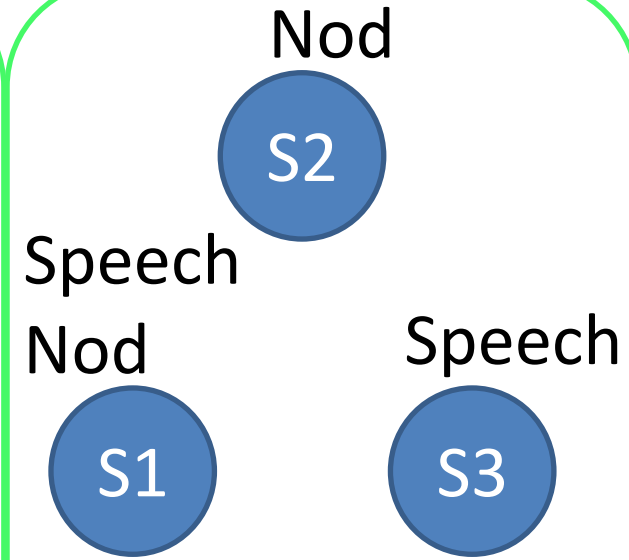
マルチモーダルイベント発見の結果 100パターン以上頻出するパターンの典型例



903パターン



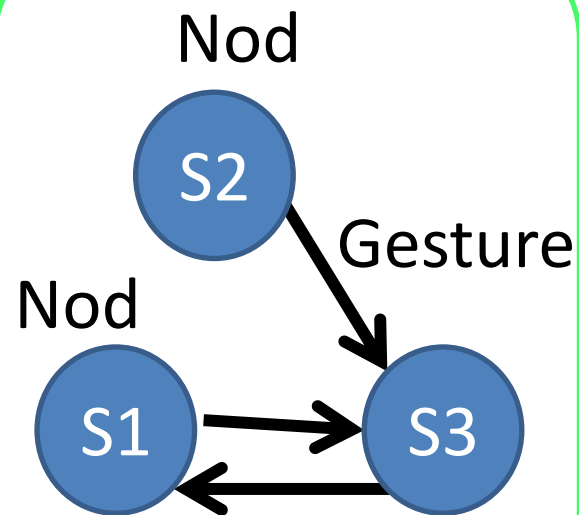
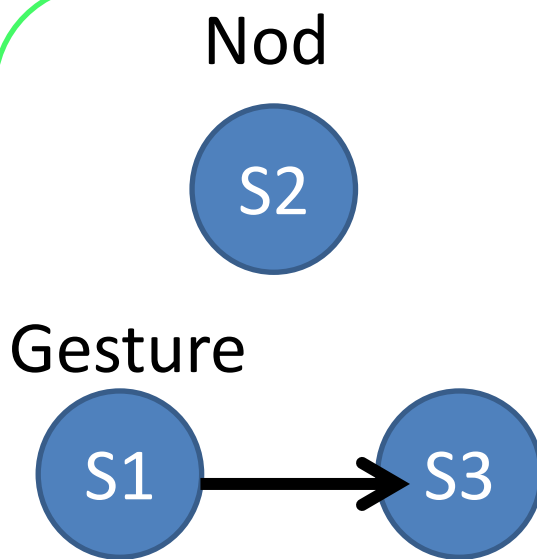
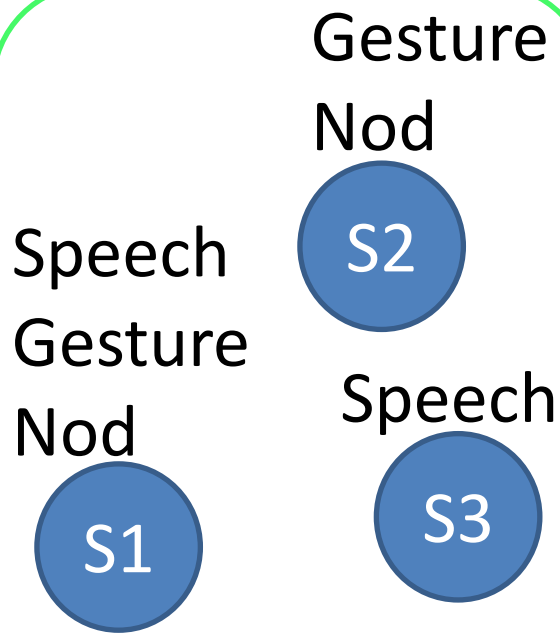
433パターン



433パターン

他にも、(S1が発話 S3がうなづく)などのパターンが頻出

マルチモーダルイベント発見の結果 ジェスチャに関連する頻出パターンの典型例



マルチモーダルイベント発見の結果 セッションごとの比較

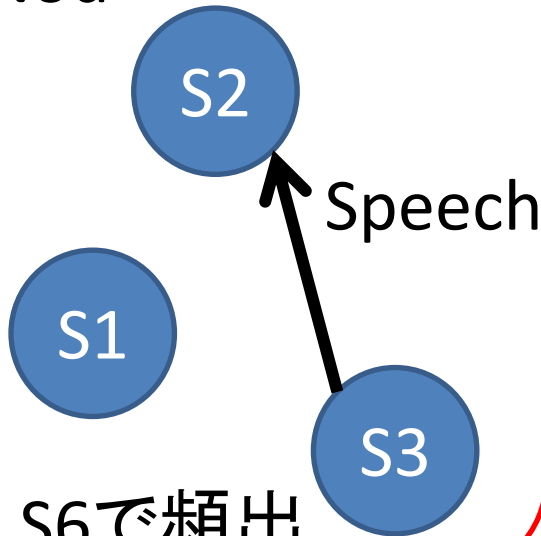
特徴的なイベントに関する各セッションごとの頻度

	S1	S2	S3	S4	S5	S6	S7	S8
聞き手の インタラクション	14	0	1	9	0	26	17	8
アドレスされてい ない話者が行う ジェスチャ	24	2	38	0	0	0	0	20
ジェスチャの引き 継ぎ	19	0	10	0	0	5	15	0

マルチモーダルイベント発見の結果 セッションごとの比較

聞き手の
インタラクション

Nod



パターンの条件

1. 聞き手S3の発話・腕部動作がon
2. 話し手の発話・腕部動作がoff
3. 話し手のいずれかは頭部縦動作on

特徴的なポイント

聞き手のインタラクションにより説明者の説明を引き出している

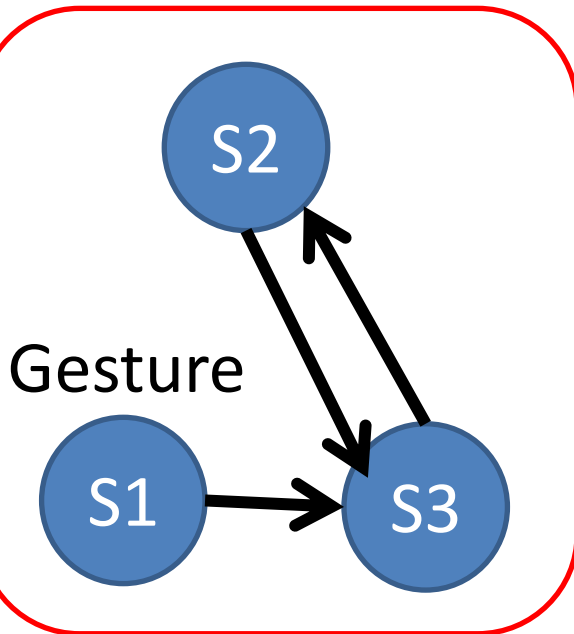
	S1	S2	S3	S4	S5	S6	S7	S8
聞き手の インタラクション	14	0	1	9	0	26	17	8

マルチモーダルイベント発見の結果 セッションごとの比較

アドレスされていない
話者が行うジェスチャ

パターンの条件

1. 聞き手S3がSxに顔向けしている
2. Sxでない話し手がジェスチャを行う



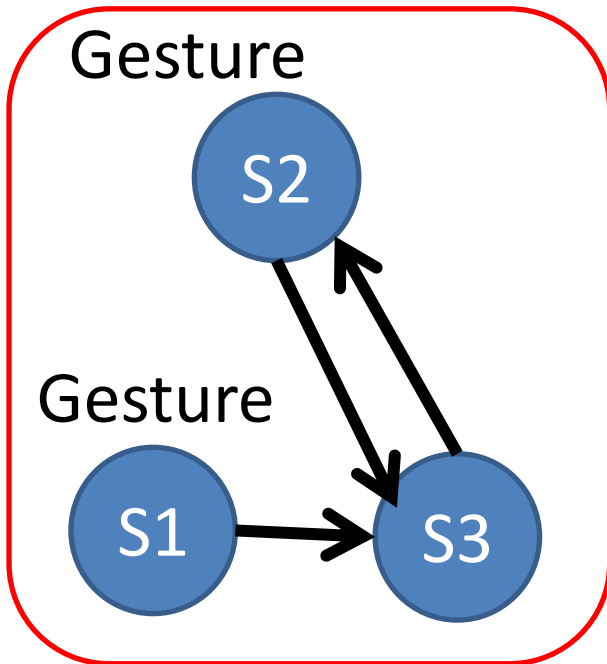
特徴的なポイント

アドレスされていないのに手が動く.

	S1	S2	S3	S4	S5	S6	S7	S8
アドレスされていない話者が行う ジェスチャ	24	2	38	0	0	0	0	20

マルチモーダルイベント発見の結果 セッションごとの比較

ジェスチャの引き継ぎ



パターンの条件

1. 説明者1のジェスチャと説明者2のジェスチャが共起する

特徴的なポイント

協調的説明（井上・他 2011）

	S1	S2	S3	S4	S5	S6	S7	S8
ジェスチャの引き継ぎ	19	0	10	0	0	5	15	0

マルチモーダルデータマイニングに関する 研究のまとめと課題

系列マイニングに基づくデータマイニング手法を提案

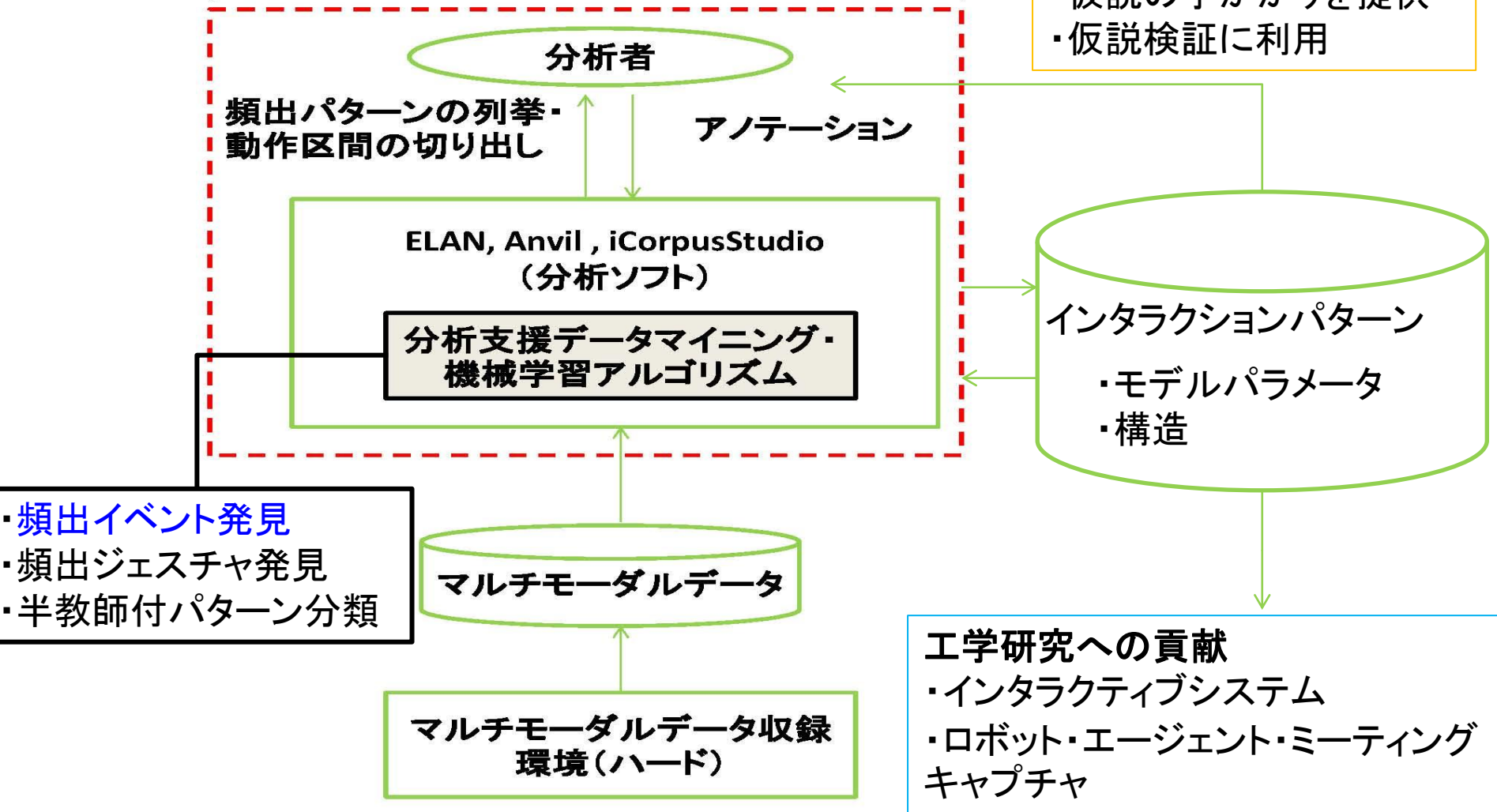
1. 腕部・頭部動作の抽出を高精度に行える手法を開発・実装
2. 部分空間パターンの抽出, 時間のずれを許容した柔軟なマッチングを行えるCMPDアルゴリズムを開発した
⇒パターンの計数や共起パターン発見を行えることを示した.

アニメーション説明課題タスクへの適用:

- 各セッションに特有のイベントを発見した
⇒ 時間順序を加味したイベント発見アルゴリズムに拡張する
- 参加者の顔向け状態を加味することで説明以外に用いられるジェスチャを抽出した
⇒ 顔向け状態を加味して, ジェスチャの機能分類を試みる

本研究のロードマップ

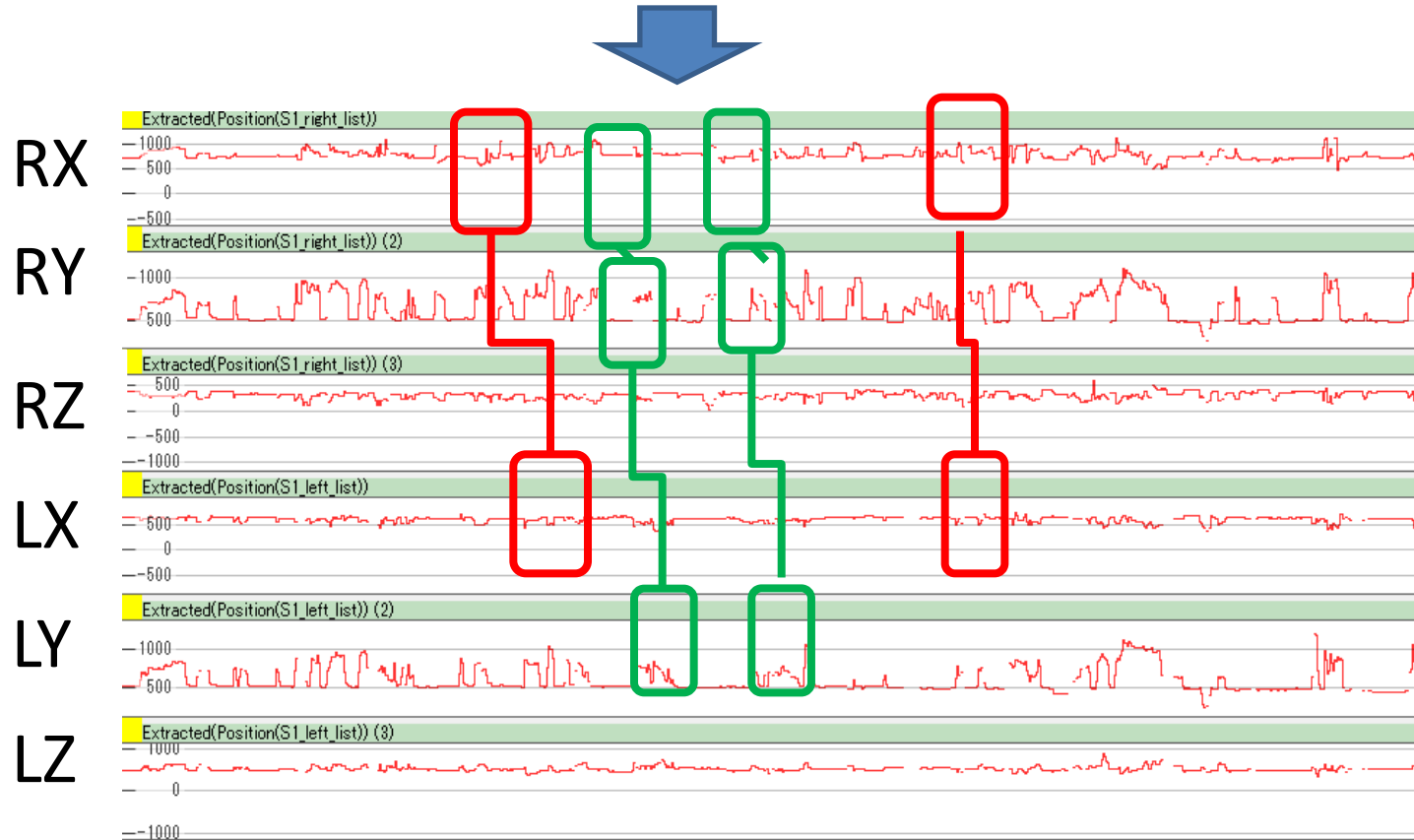
本研究の枠組み (分析援助のために機械学習が存在)



- その他のデータマイニングツール
 - 頻出ジェスチャパターンの発見
 - 少数アノテーションを利用した
頻出ジェスチャパターンの分類

頻出ジェスチャパターンの発見

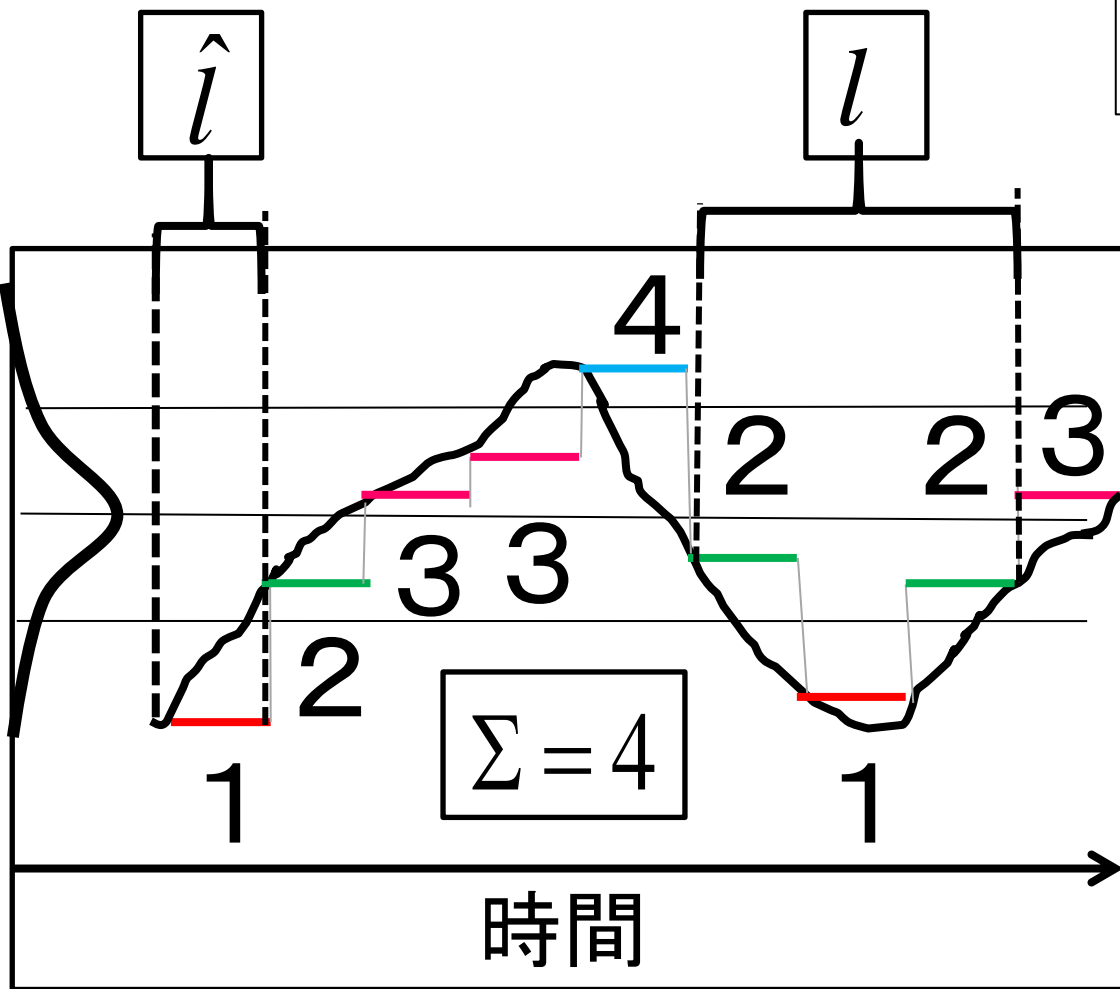
時系列データマイニングによるパターン抽出



頻出ジェスチャパターン

一次元時系列データからのパターン発見

Random Projection



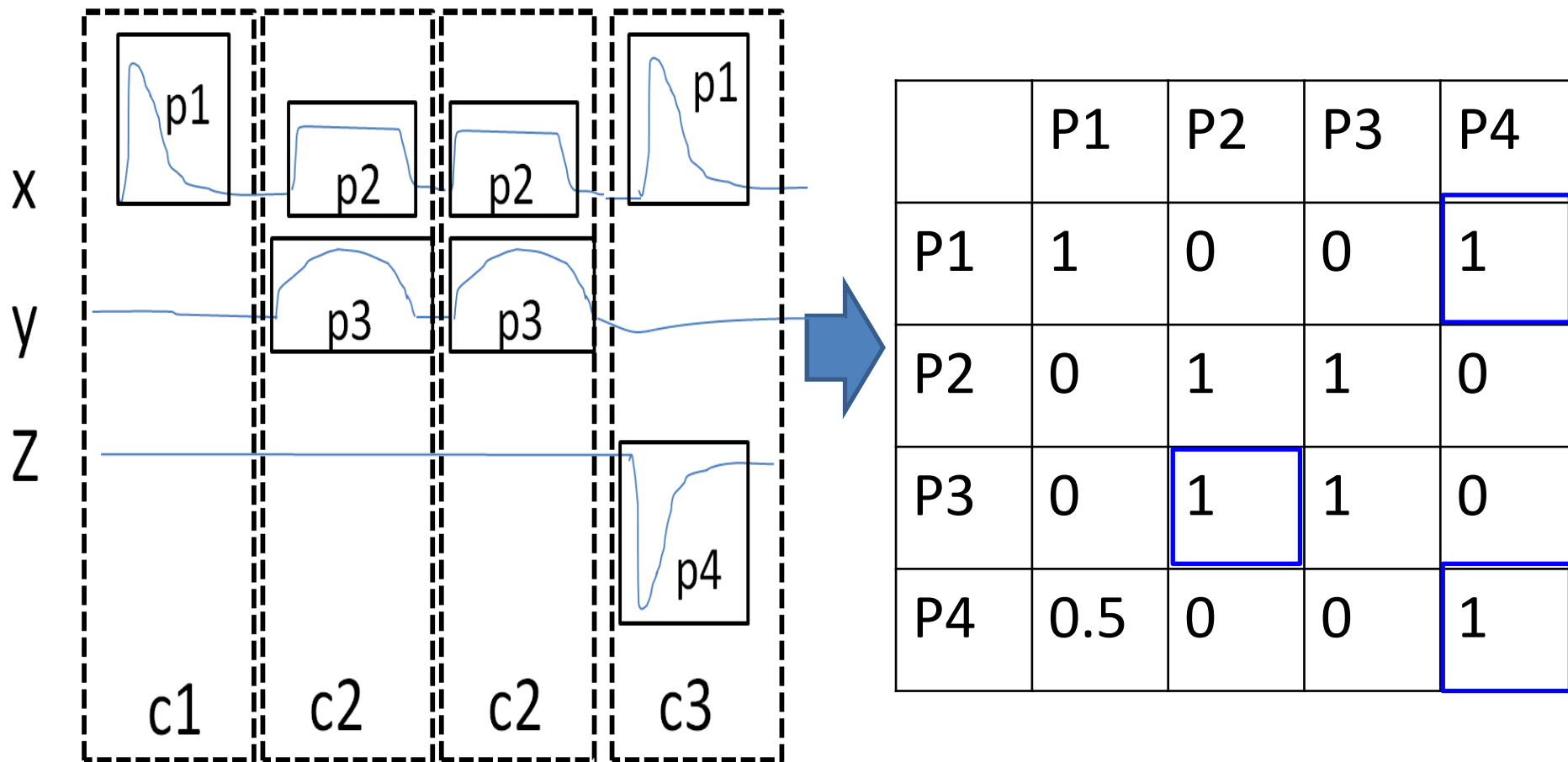
1	2	3
2	3	3
3	3	4
3	4	2
4	2	1
2	1	2
1	2	3

SAX

123342123

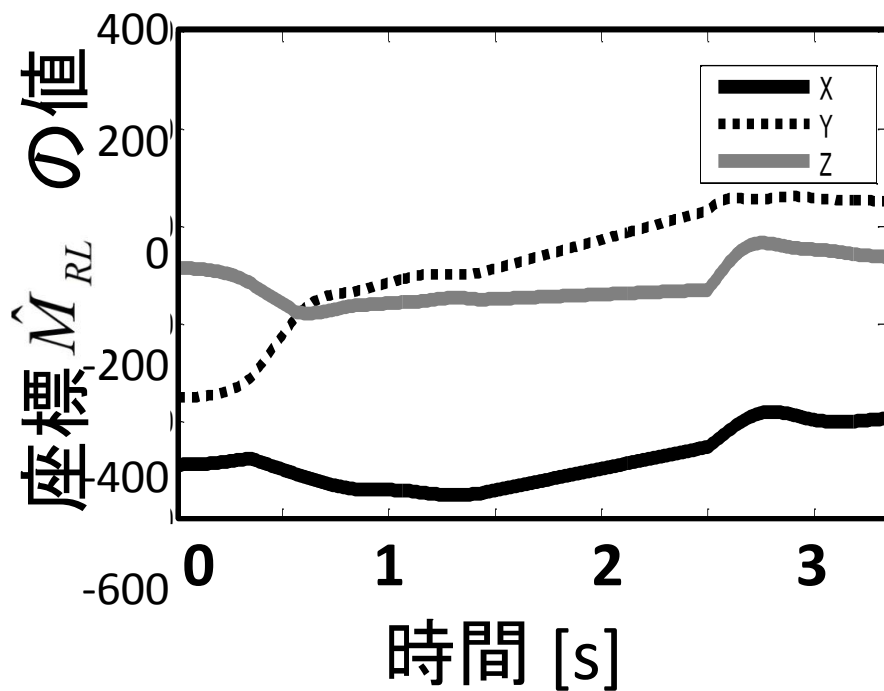
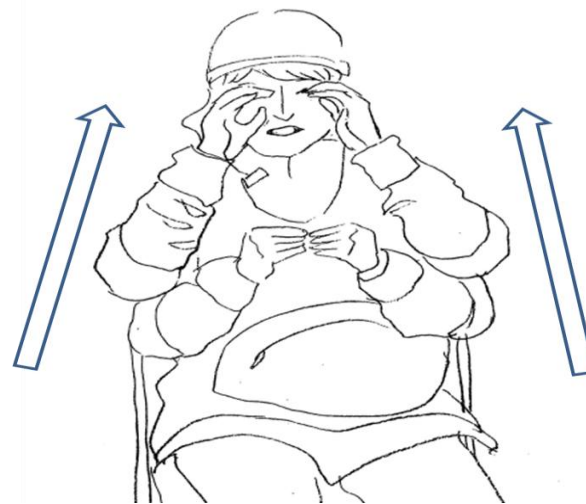
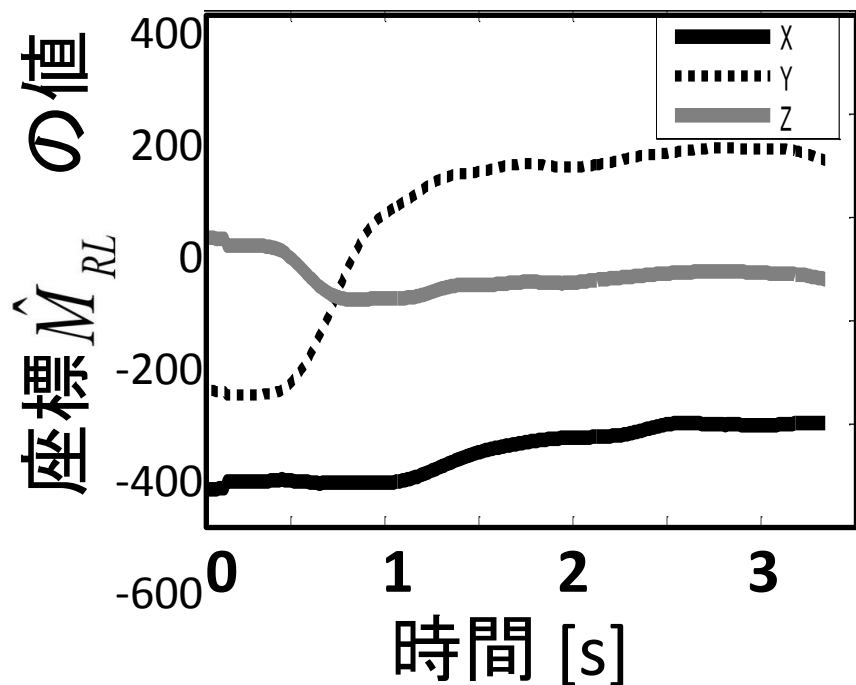
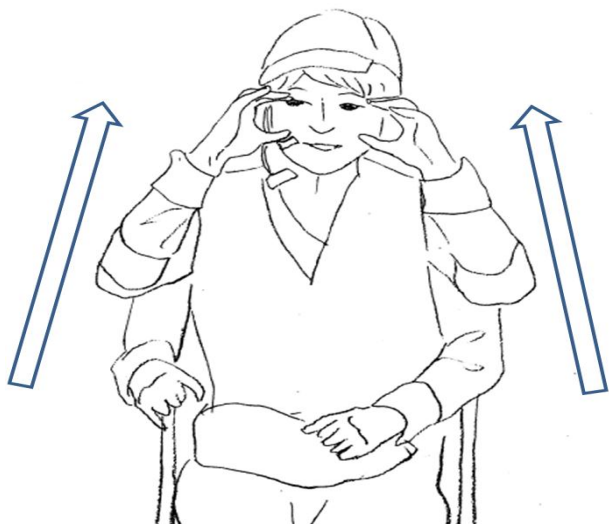
(1, 2, 3)がパターンとして抽出される

各次元で発見されたパターンの統合手法

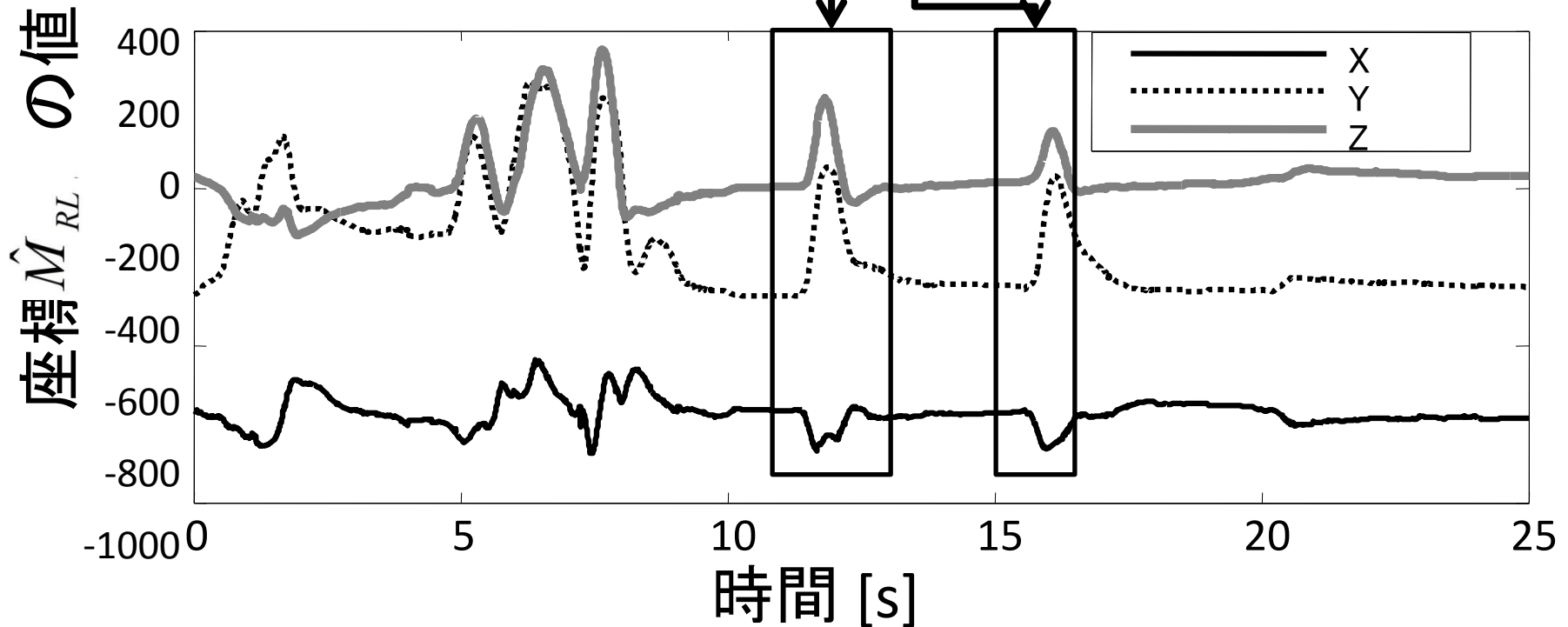
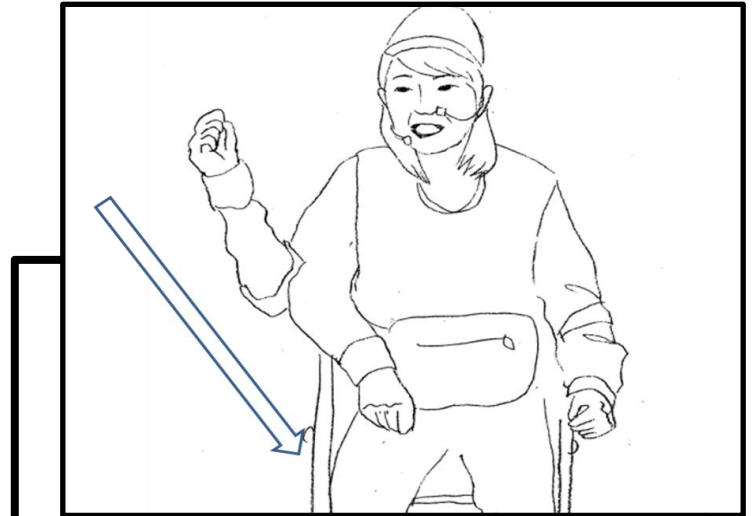
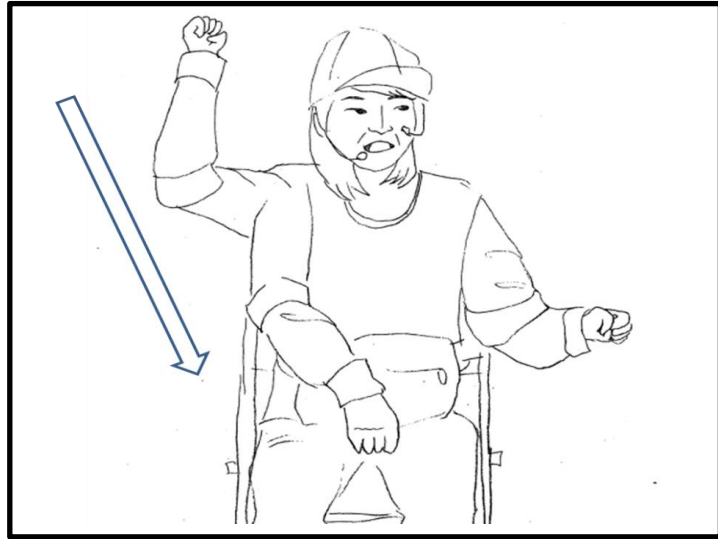


1. 各パターンの共起度を計算
2. オーバーラップすれば共起とみなす

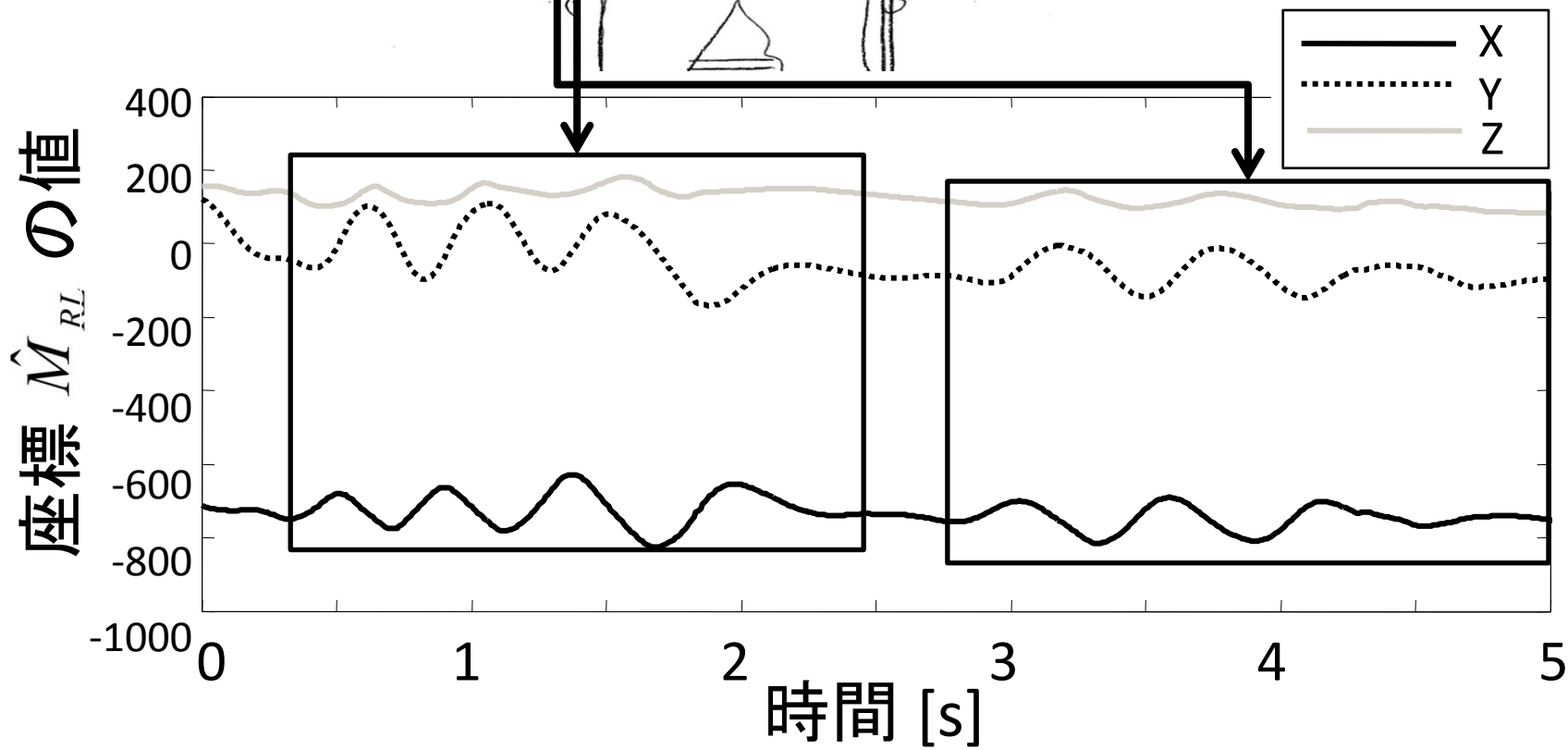
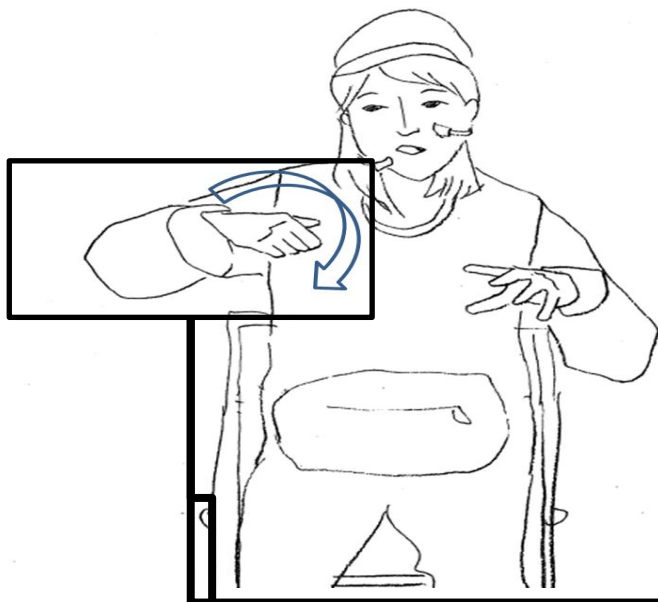
(抽出例) 双眼鏡を覗くしぐさのジェスチャ



(抽出例)おばあさんが傘を振り下ろすジェスチャ

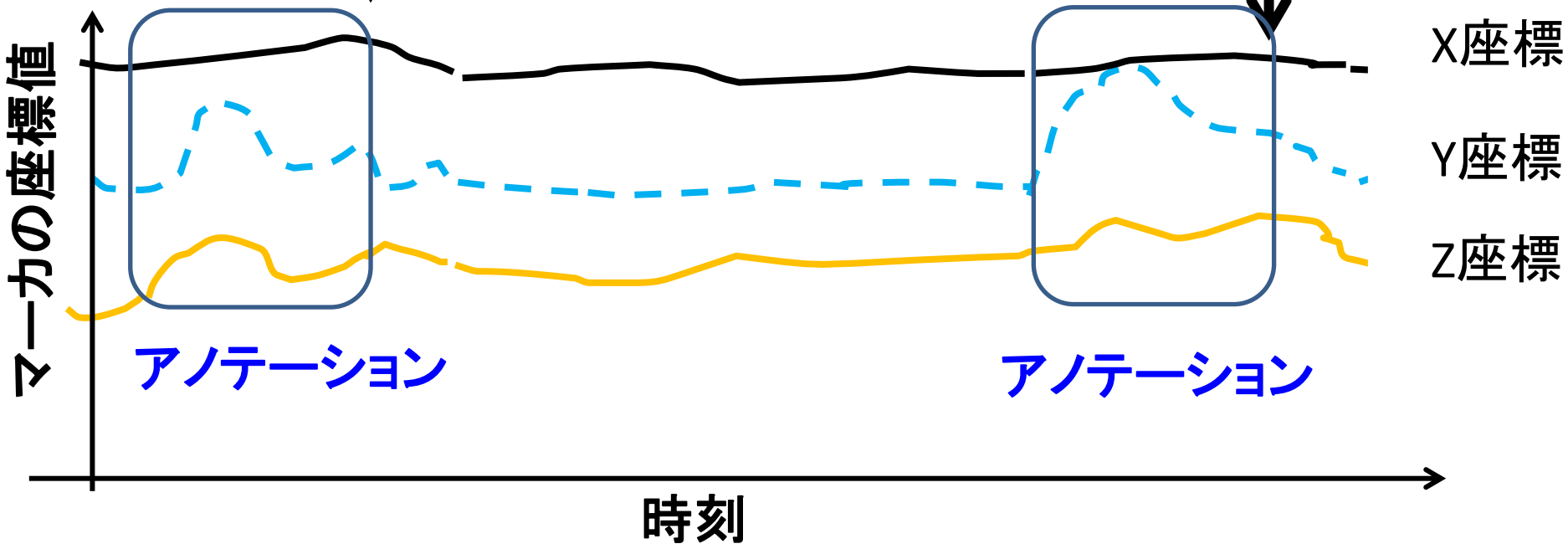


(抽出例)手回しオルガンを弾いているシーンを示すジェスチャ



- その他のデータマイニングツール
 - 頻出ジェスチャパターンの発見
 - 少数アノテーションを利用した
頻出ジェスチャパターンの分類

分析者によるアノテーションからの学習



少量のアノテーションデータを利用してデータを分類する半教師付き学習手法

半教師付き学習とは少量のアノテーションされたデータと大量の未アノテーションデータを用いて学習し、全てのデータにアノテーションを付与する (Belkin & Niyogi)

- メリット:
 - 専門家による少量のアノテーションデータを学習に利用
 - アノテーションの手間を軽減可能

Graph-based SSC approach [Zhou 2003,Zhu 2003,Bengio 2006]

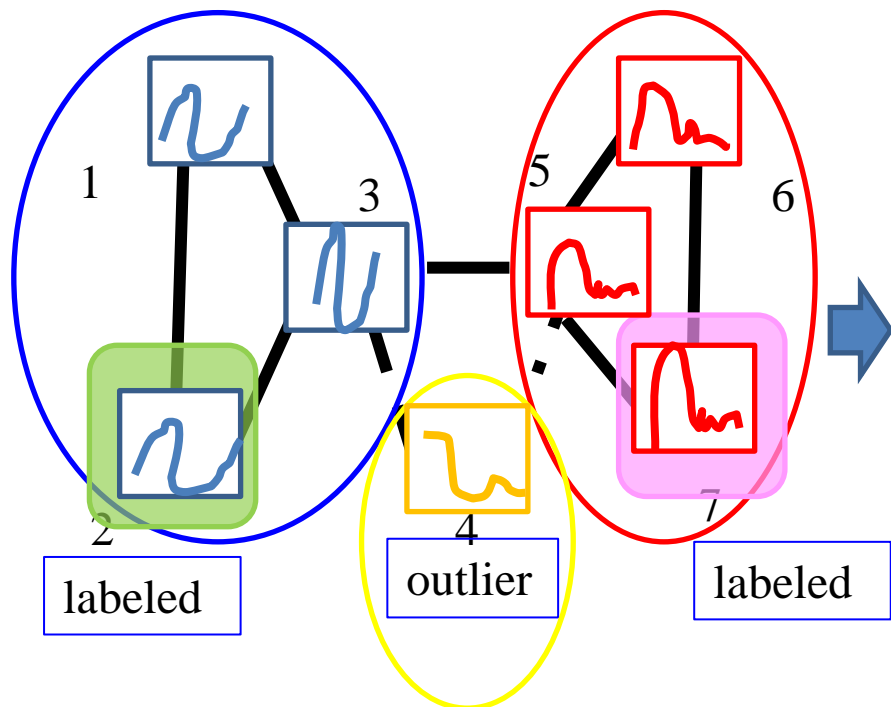
- 多様体仮説に基づいている
- 多クラス問題に簡易に対応可能



提案システムの半教師付き分類手法にグラフに基づく手法を採用

グラフに基づく半教師付き学習

Undirected graph of samples



Affinity/ Similarity Matrix

$$W = \begin{pmatrix} 1 & 0.8 & 0.6 & 0 & 0 & 0 & 0 \\ 0.8 & 1 & 0.7 & 0 & 0 & 0 & 0 \\ 0.6 & 0.7 & 1 & 0.01 & 0 & 0 & 0 \\ 0 & 0 & 0.01 & 1 & 0.001 & 0 & 0 \\ 0 & 0 & 0 & 0.001 & 1 & 0.8 & 0.9 \\ 0 & 0 & 0 & 0 & 0.8 & 1 & 0.95 \\ 0 & 0 & 0 & 0 & 0.9 & 0.95 & 1 \end{pmatrix}$$

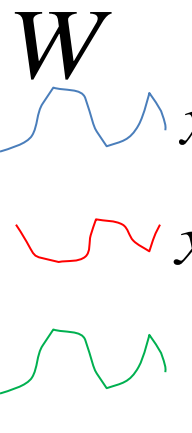
Label Propagation

Learning or optimize W

半教師付き学習アルゴリズム

(Okada, Nishida ICANN 2010, Wei Liu et al CVPR 2009)

Adjacency (Similarity) Matrix W


$$W = \begin{matrix} & x_1 & x_2 & \cdots \\ x_1 & \begin{pmatrix} 1 & w_{12} & \cdots \\ w_{21} & 1 & w_{23} \\ \vdots & w_{32} & \ddots \end{pmatrix} \\ x_2 & \\ \vdots & \end{matrix}$$

Labeled data X_l
(x : pattern, y : label)

$$X_l = \{(x_1, y_1), \dots, (x_l, y_l)\}$$

Unlabeled data X_U

$$X_U = \{x_{l+1}, x_{l+2}, \dots, x_N\}$$

・DTWとGaussian Kernelを用いて
類似度行列を作成

Construction Graph

$$w_{i,j} = \begin{cases} 0, & \text{if } y_i \neq y_j \\ d_M(\hat{x}_i, \hat{x}_j), & \text{otherwise} \end{cases}$$

RMGT

Learning Adjacency Matrix

Label Propagation

半教師付き学習アルゴリズム (類似度行列の学習)

類似度行列の学習

$$W_{i,j} = \begin{cases} 0, & \text{if } y_i \neq y_j \\ d_M(\hat{x}_i, \hat{x}_j), & \text{otherwise} \end{cases}$$



$$\min \mathcal{G}(\mathbf{W}) = \frac{1}{2} \|\mathbf{W} - \mathbf{W}^0\|_F^2$$

$$\text{s.t.} \quad \sum_{(i,j) \in \mathcal{T}} W_{ij} = 0$$

$$\mathbf{W}\mathbf{1} = \mathbf{1}, \mathbf{W} = \mathbf{W}^T, \mathbf{W} \geq 0$$

Wの更新

$$\mathbf{W} = \mathcal{P}(\mathbf{W}^0, \mathbf{T}) = \mathbf{W}^0 - \left(t^0 + \frac{2\mathbf{1}^T \mathbf{T} \boldsymbol{\mu}^0}{|\mathcal{T}|} \right) \mathbf{T} \\ + \boldsymbol{\mu}^0 \mathbf{1}^T + \mathbf{1} \boldsymbol{\mu}^{0T},$$

$$t^m = \sum_{(i,j) \in \mathcal{T}} \mathbf{W}_{ij}^m / |\mathcal{T}|, \boldsymbol{\mu}^m = \hat{\mathbf{T}}(\mathbf{1} - \mathbf{W}^m \mathbf{1} + t^m \mathbf{T} \mathbf{1})$$

半教師付き学習アルゴリズム

Adjacency (Similarity Matrix W)

$$W = \begin{matrix} & x_1 & x_2 & \cdots \\ \begin{matrix} x_1 \\ x_2 \\ \vdots \end{matrix} & \begin{pmatrix} 1 & w_{12} & \cdots \\ w_{21} & 1 & w_{23} \\ \vdots & w_{32} & \ddots \end{pmatrix} \end{matrix}$$

Labeled data X_l
(x : pattern, y : label)

$$X_l = \{(x_1, y_1), \dots, (x_l, y_l)\}$$

Unlabeled data X_U

$$X_U = \{x_{l+1}, x_{l+2}, \dots, x_N\}$$

Normalized Cutに基づき
計算された固有ベクトルから
類似度行列を作成

Construction Graph

$$w_{i,j} = \begin{cases} 0, & \text{if } y_i \neq y_j \\ d_M(\hat{x}_i, \hat{x}_j), & \text{otherwise} \end{cases}$$

RMGT

Learning Adjacency Matrix

Label Propagation

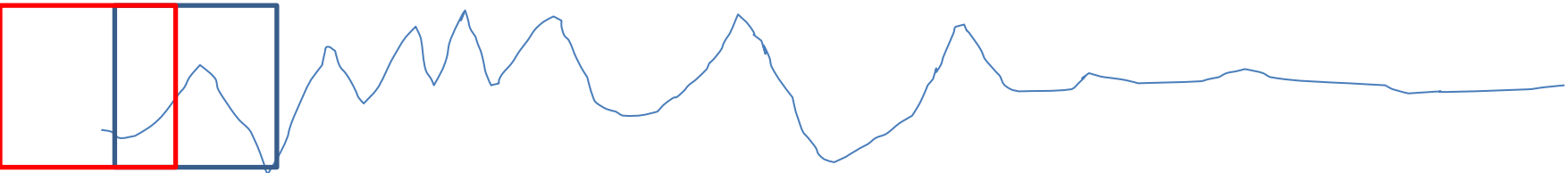
関連研究: KNN法に基づくテンプレートマッチング

シンプルなKNN探索

1 取り出したいジェスチャパターン(クエリ)を準備





2 時系列データをクエリでマッチング



シンプルなKNN探索のデメリット

未アノテーションデータを学習に利用できない。

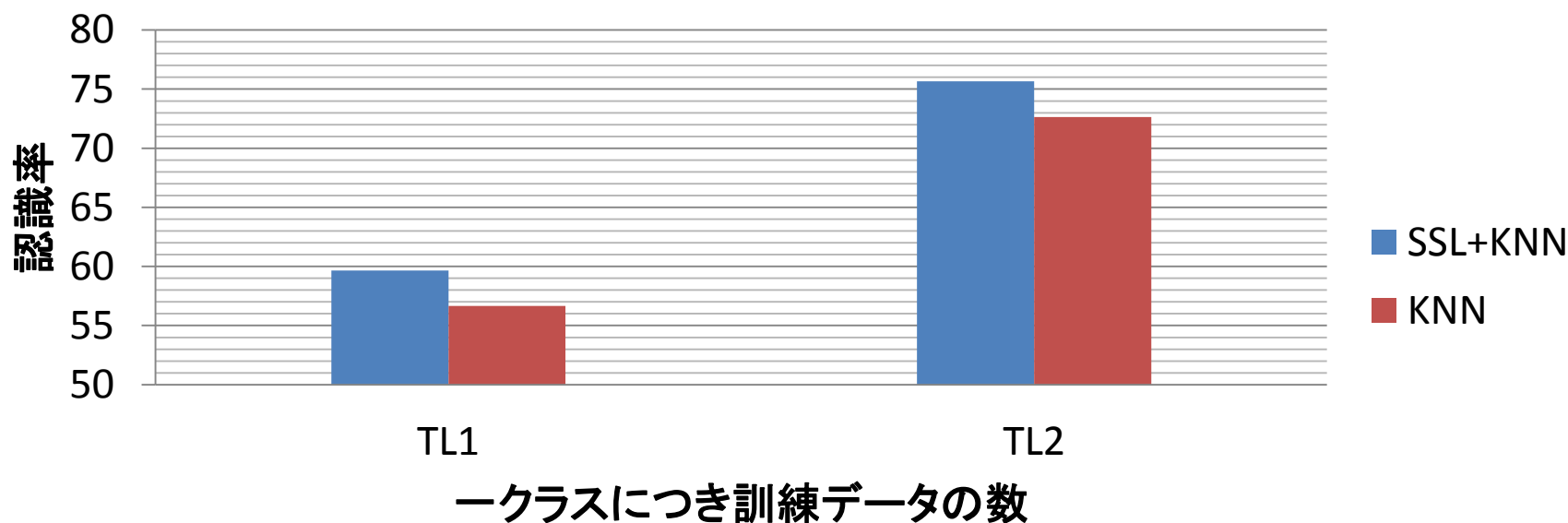
- ・  と  の違いを利用する
- ・ 未アノテーションデータを利用する (半教師付き学習)

アノテーション精度の比較結果

- UCI machine learning dataset (Aust.Gesture)
- ラベルデータを入れ替えて20回の実験



- KNN法と提案法のアノテーション付与の精度を比較



訓練データが1個の場合、最大5.3%(136個)分類精度を改善

参考文献

1. Carletta, Jean et al , The AMI meeting corpus: A pre-announcement. In Second International Workshop on Machine Learning for Multimodal Interaction (MLMI 2005), Vol.3869 of Lecture Notes in Computer Science, pp.28-39, 2005
2. Shen, Lei, Rose,. VACE multimodal meeting corpus. In Second International Workshop on Machine Learning for Multimodal Interaction (MLMI 2005), Vol.3869 of Lecture Notes in Computer , 2005
3. Chiu, Bill, Keogh, Eamonn Lonardi, Stefano. Probabilistic discovery of time series motifs. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '03). ACM, pp.493-498, 2003
4. Kazuhiro, Otsuka Multimodal Conversation Scene Analysis for Understanding People's Communicative Behaviors in Face-to-face Meetings. 14th International Conference on Human-Computer Interaction (HCI International 2011), pp.171-179, 2011
5. Waibel, Alexander, & Stiefelhagen, Rainer (Ed.) , Computers in the Human Interaction Loop. Springer. , 2009
6. Daniel Gatica-Perez , Automatic nonverbal analysis of social interaction in small groups: A review, Image and Vision Computing , pp.1-13, 2009
7. 中田篤志・角康之・西田豊明 非言語行動の出現パターンによる会話構造抽出. 電子情報通信学会論文誌, Vol.J94-D, No.1, pp.113-123 , 2010
8. 坊農真弓・高梨克也 (編) 多人数インタラクションの分析手法 オーム社, 2009

当発表に関連する発表者の研究成果

1. 岡田将吾、坊農真弓、角康之、高梨克也、会話インタラクションにおけるジェスチャの量的分析を支援する時系列データマイニング手法の開発、社会言語科学 2012年 9月
2. 岡田 将吾、西田 豊明、“自己増殖型ニューラルネットワークを用いた時系列データの追加学習型クラスタリング”, 日本神経回路学会論文誌 Vol.17, No.4, 174-186
3. Shogo Okada, Osamu Hasegawa, Toyoaki Nishida, "Machine Learning Approaches for Time-series Data Based on Self-Organizing Incremental Neural Network", International Conference on Artificial Neural Networks (ICANN2010)
4. Shogo Okada, Satoshi Ishibashi and Toyoaki Nishida: "On-line Unsupervised Segmentation for Multidimensional Time-series Data and Application to Spatiotemporal Gesture data", IEA/AIE2010
5. Shogo Okada, Toyoaki Nishida, "Multi Class Semi-supervised Classification with Graph Construction Based on Adaptive Metric Learning", International Conference on Artificial Neural Networks (ICANN2010), 2010

紙面の都合ですべての参考文献を掲載しておりません。
当研究に関するその他の関連研究のお問い合わせは
okada@ntt.dis.titech.ac.jpまでお願いします。