わが国におけるマルチモーダル対話記述標準化
―現状と将来への期待―

荒木雅弘
京都工芸繊維大学

発表の構成

- ITSCJ試行標準について
  - W3C MMI WGとの比較
- 活動　フェーズ1　報告
  - 階層型アーキテクチャに至るまで
- 活動　フェーズ2　報告
  - ユースケース検討から各階層の仕様まで
- 今後の予定

ITSCJ試行標準について

- 情報処理学会
  情報規格調査会
  「音声入出力インタフェース委員会」
  (2007年4月～)
- ミッション
  - 「試行標準」として仕様情報をWebで公開
    - 第一版を近日公開予定
  - 意見を募って拡張・改良を行い、標準規格へ

ITSCJ試行標準について

- 検討内容
  - MMIシステムのアーキテクチャ
  - 各コンポーネントの要求仕様
- 目指すところ
  - MMIシステム実装のガイドライン作成
  - 開発用フレームワークへ
  - W3Cに提案し、国際標準へ

我々の活動の目標

1. 実用システムにも**研究開発**にも利用できるアーキテクチャを確立

   ⇔ W3C: モバイル・アクセシビリティの確保を中心に据えた**現実的**な視点

2. 実装事例を通じて妥当性を検討

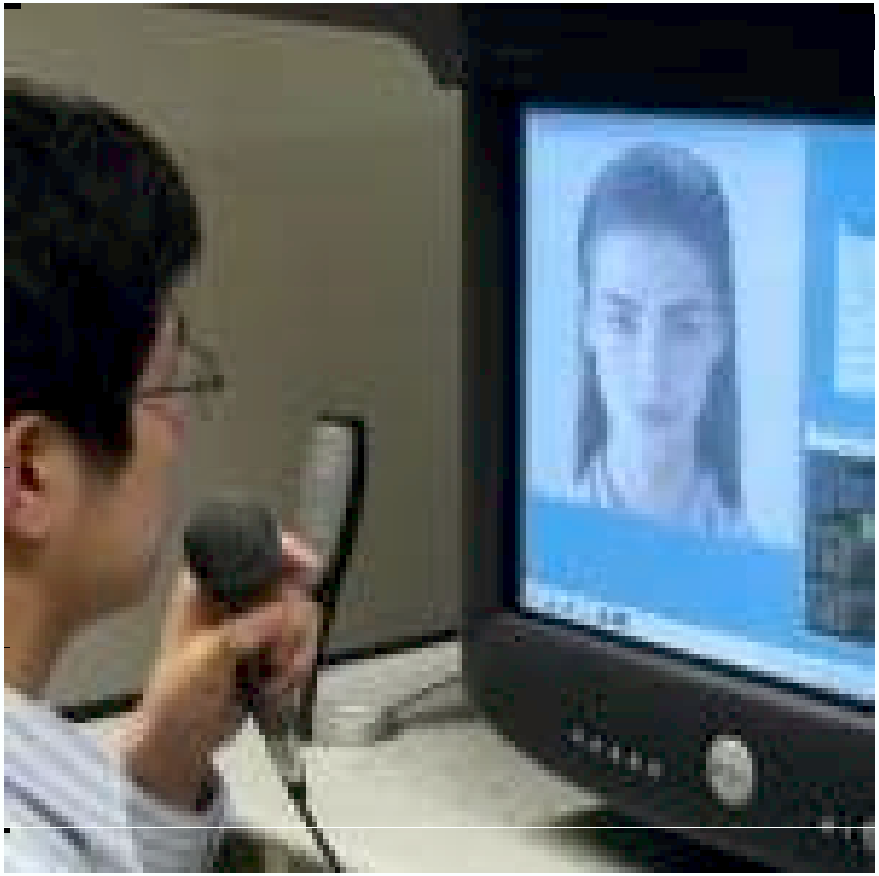
   ⇒ Galatea Toolkit などで実装

3. 開発用フレームワークとしてリリース

   ⇒ 実績をアピールし、国際標準へ

Galatea Toolkit(1)

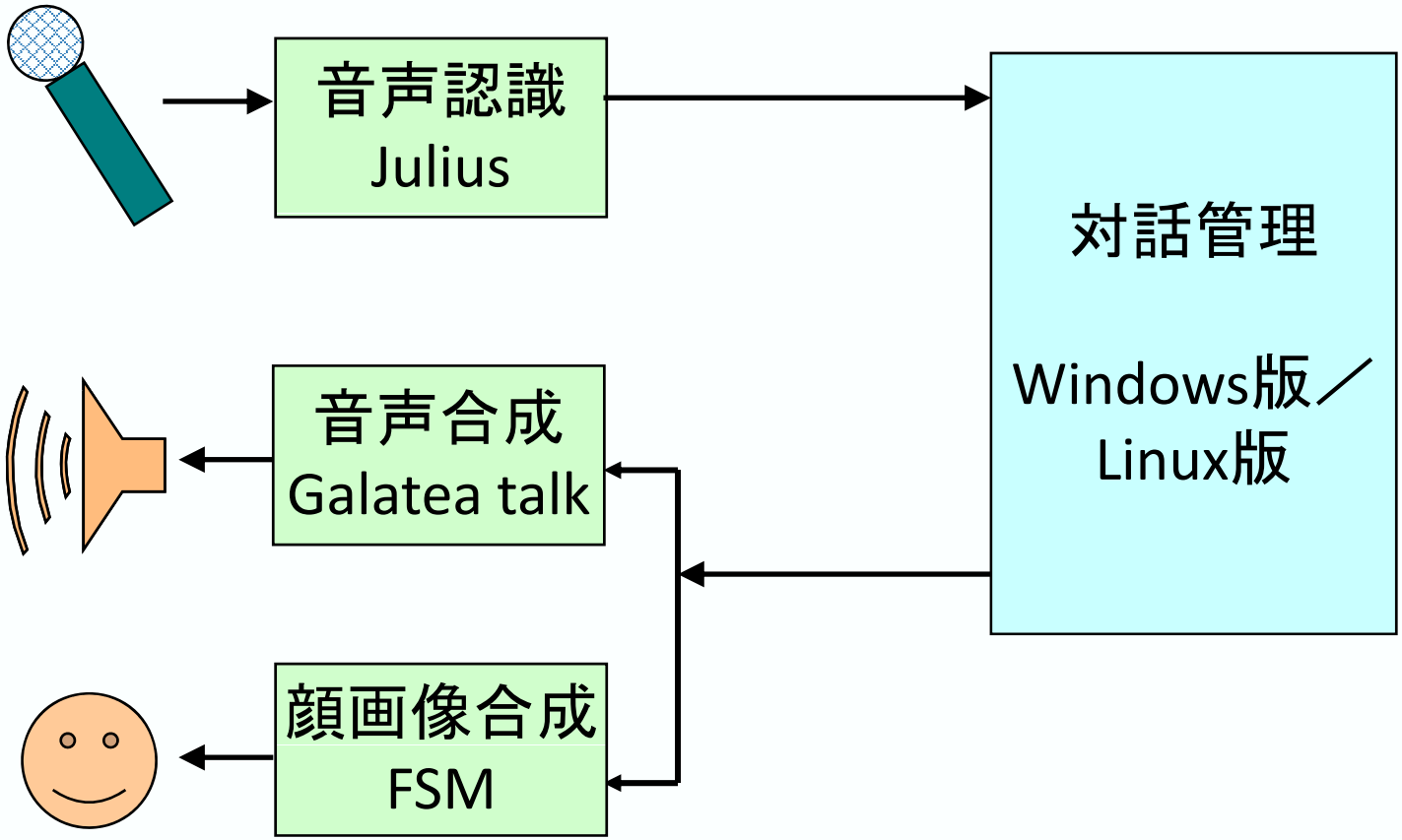
- MMIシステム開発のプラットフォーム
  - 音声認識
  - 音声合成
  - 顔画像合成



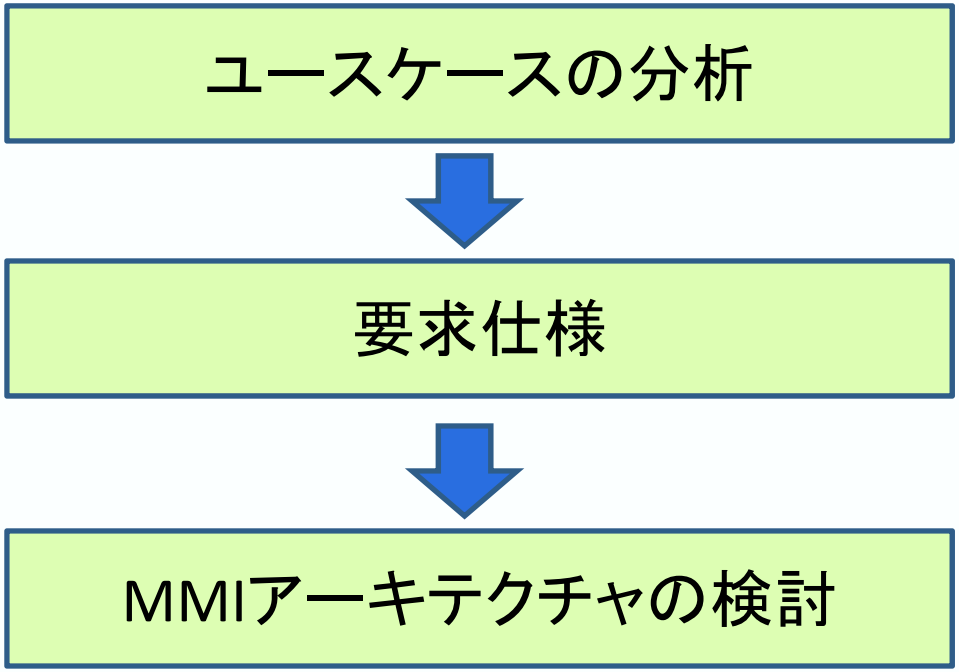
- オープンソース

http://sourceforge.jp/projects/galatea/

Galatea Toolkit(2)



検討の手順
第1フェーズ

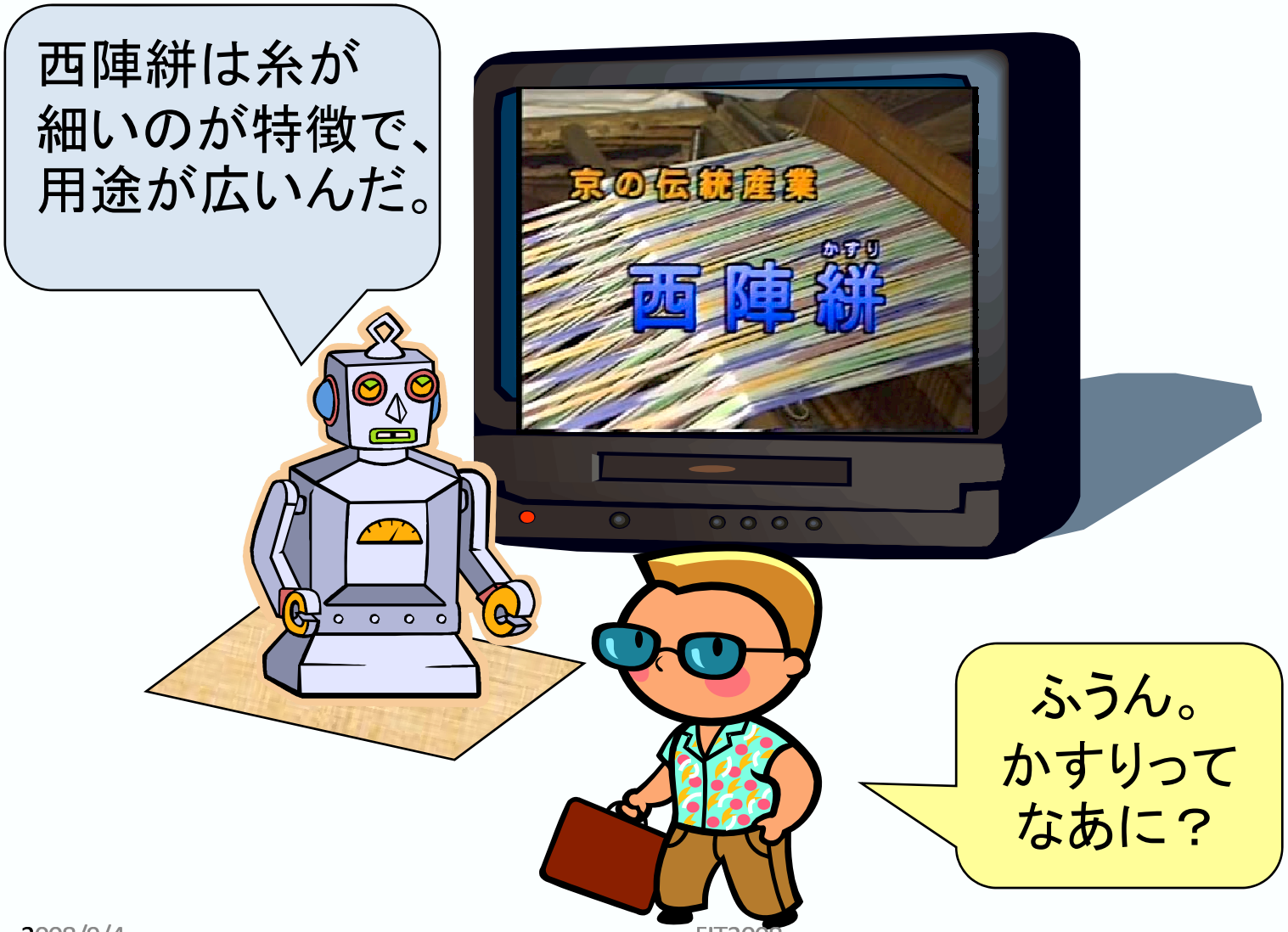


検討結果は http://www.astem.or.jp/istc/ にて公開中

ユースケース分析

	内容	入出力モダリティ	概要
a	オンラインショッピング	入力：マウス，音声 出力：ディスプレイ，音声，エージェント	音声インタフェースを備えたPC上でのオンラインショッピング．ユーザに応じたショッピングコンテンツの選択や，多様なモダリティを組み合わせたユーザ入力が扱われる．
b	音声によるディレクトリ検索	入力：マウス，音声 出力：ディスプレイ，音声	音声による飲食店の検索．主にスロットフィリング型の対話を扱っており，フィールド値に応じた対話進行や，フィールド値の動的変更等が行われる．
c	サイト検索	入力：マウス，音声，キー 出力：ディスプレイ，音声	PCあるいは携帯電話を端末とした，音声・キーによるサイト検索．
d	ロボットとの対話	入力：画像，音声，センサ 出力：ディスプレイ，音声	カメラやセンサ，ディスプレイを備えたロボットとの対話．画像によるユーザの認識や，接触によるインタラクション，webページの表示やビデオコンテンツの再生を行う．
e	対話エージェントとの交渉	入力：音声 出力：音声，エージェント	歯科での予約受付を想定した対話エージェントとの交渉．ユーザ発話中のシステムの割り込み，ユーザ動作や状況の理解を行う．
f	音声情報案内システム	入力：マウス，音声 出力：ディスプレイ，音声	施設内の設備案内，交通情報，ニュース等の情報提供を行う情報キオスク．年少のユーザによる予想外の行動（罵声や意味の無い音声，マイクを叩いた音など）にも対応する．
g	エリアガイド	入力：ペン，音声 出力：ディスプレイ，音声	エリアガイドに関するユースケースである．ペンジェスチャによるユーザ入力を受け付ける．
h	カーナビ目的地設定	入力：タッチ，音声，PTT 出力：ディスプレイ，音声	プッシュトゥトーク機能を備えた音声認識を行うカーナビゲーションシステム．

ユースケースの例
ロボットとの対話



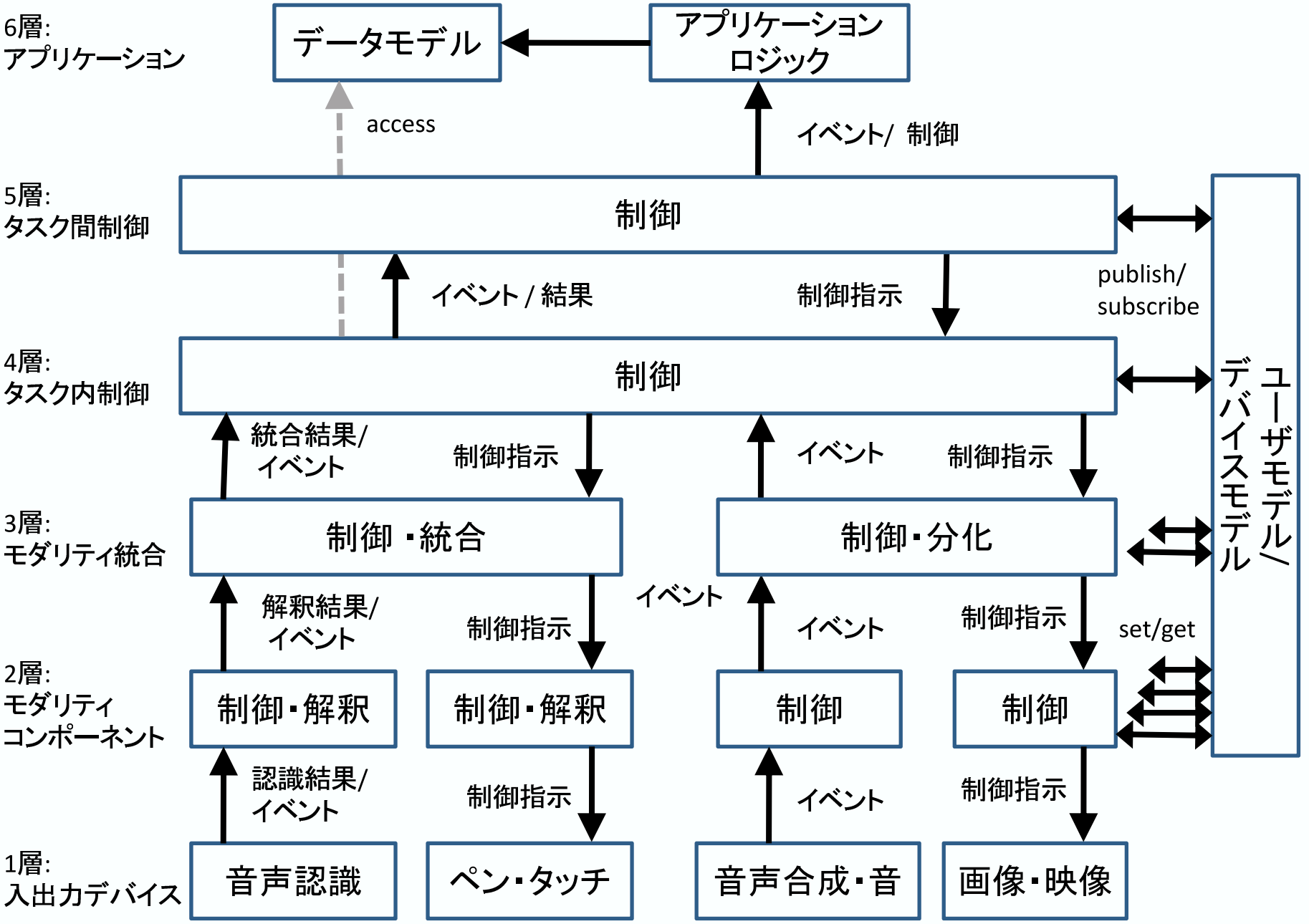
要求仕様

1. 一般的な要求
2. 入力モダリティに関する要求
3. 出力モダリティに関する要求
4. アーキテクチャ、統合、同期について
5. 実行時と配置
6. 対話制御について
7. フォーム・フィールドのハンドリング
8. アプリケーション・外部モジュールとの連携
9. ユーザ情報・環境情報
10. 開発者の視点から見た機能

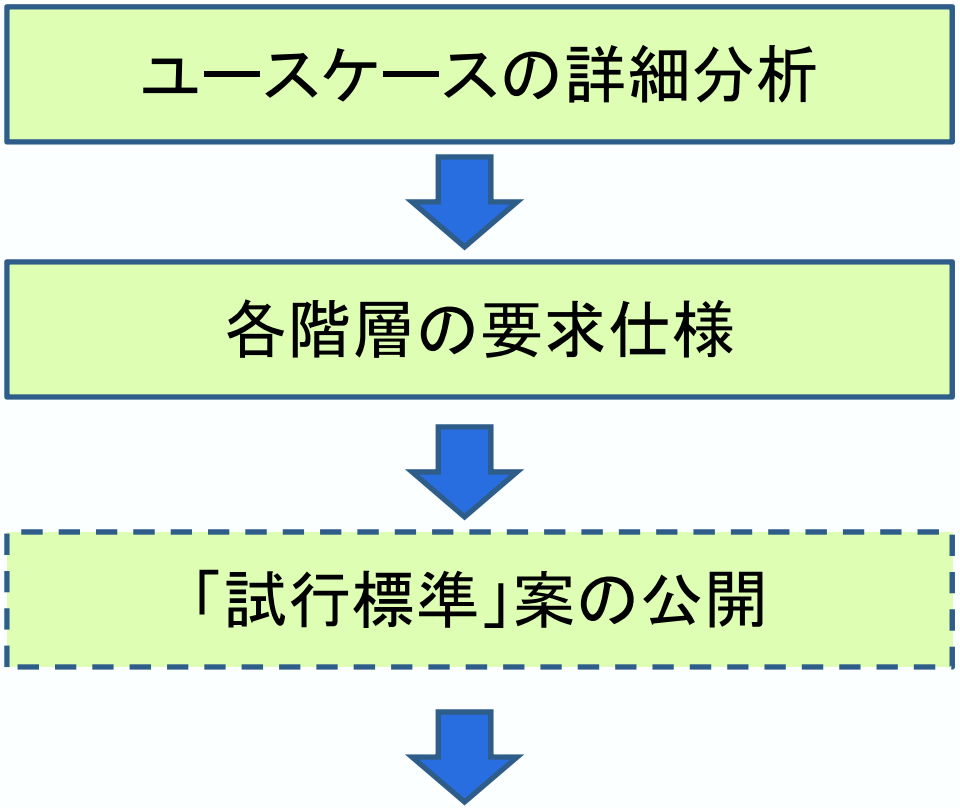
W3Cと
ほぼ共通

拡張

階層的アーキテクチャ

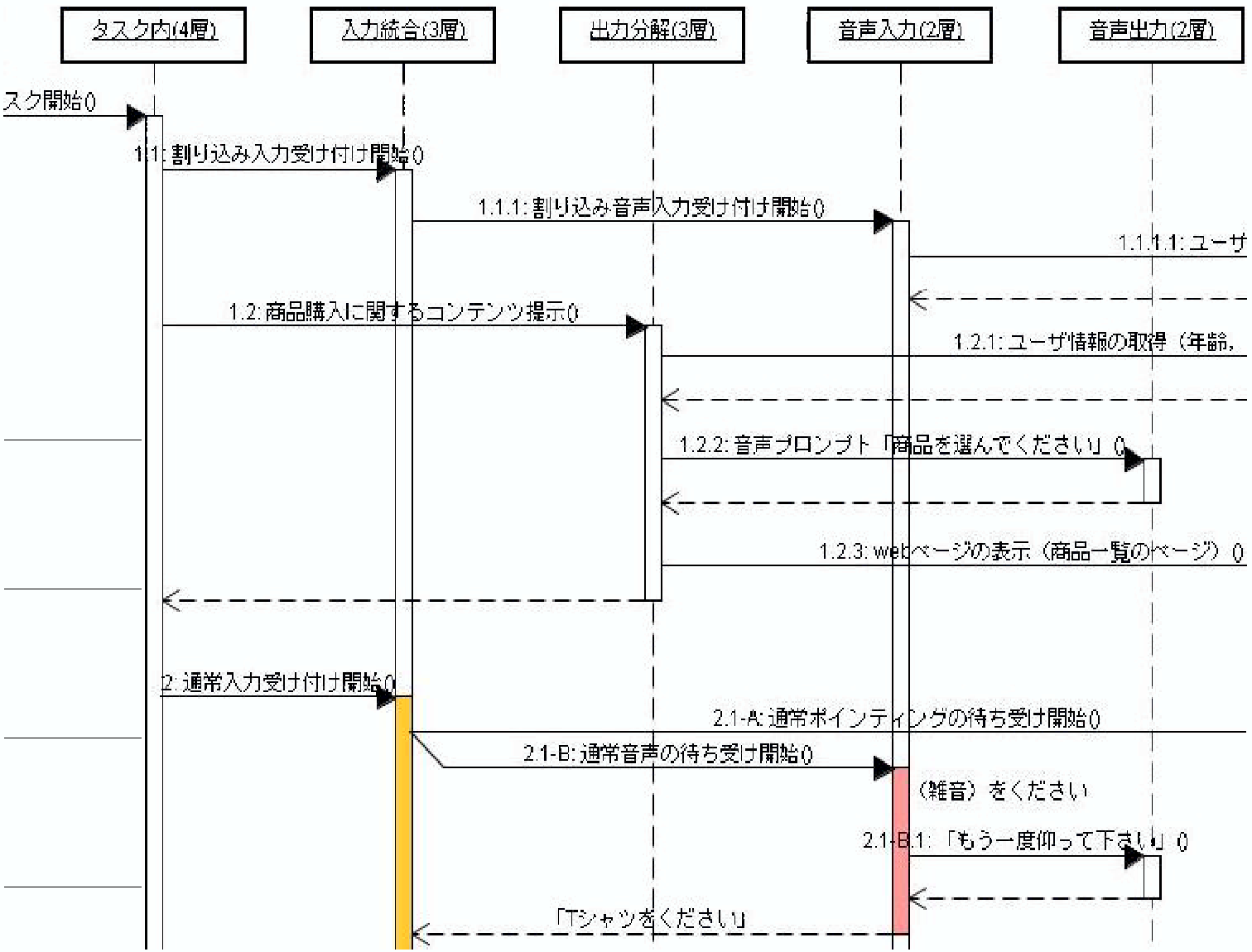

6層:
アプリケーション
データモデル
アプリケーション
ロジック
access
イベント/ 制御
5層:
タスク間制御
制御
publish/
subscribe
イベント / 結果
制御指示
4層:
タスク内制御
制御
ユーザモデル/
デバイスモデル
統合結果/
イベント
制御指示
イベント
制御指示
3層:
モダリティ統合
制御・統合
制御・分化
イベント
解釈結果/
イベント
制御指示
イベント
制御指示
set/get
2層:
モダリティ
コンポーネント
制御・解釈
制御・解釈
制御
制御
認識結果/
イベント
制御指示
イベント
制御指示
1層:
入出力デバイス
音声認識
ペン・タッチ
音声合成・音
画像・映像

検討の手順
第2フェーズ



参照実装・オープンソースフレームワークの公開へ

ユースケースの詳細分析



1層：入出力デバイス

- 機能
  - 単独モダリティの認識・合成モジュール
- 入力モジュール
  - 入力：(外部から) 信号
    (2層から)　認識処理に用いる情報
  - 出力：(2層へ) 認識結果
  - 事例：Julius, タッチ入力, 顔検出, ...
- 出力モジュール
  - 入力：(2層から) 出力内容
  - 出力：(外部へ) 信号
  - 事例：Galatea talk, FSM, Webブラウザ, ...

2層：モダリティコンポーネント

- 機能
  - 1層の多様性を吸収するラッパー

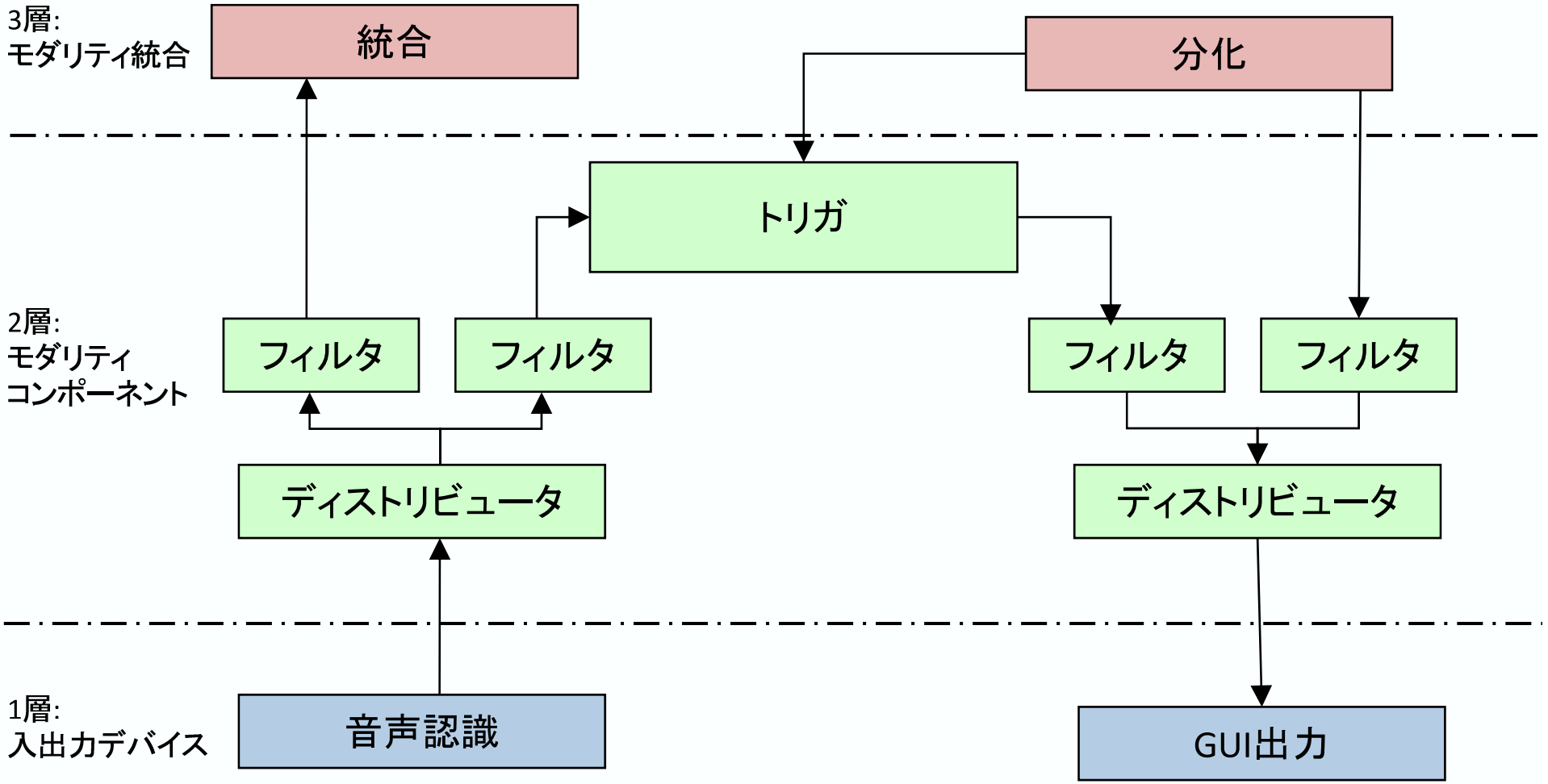
    例）音声認識コンポーネントの振る舞いの統一

    認識文法：SRGS　意味構成規則: SISR　認識結果: EMMA
  - 複数の1層の機能をまとめて、単独機能に見せる

    例) リップシンク機能付き音声合成
    音声認識の即時結果表示

2層：モダリティコンポーネント

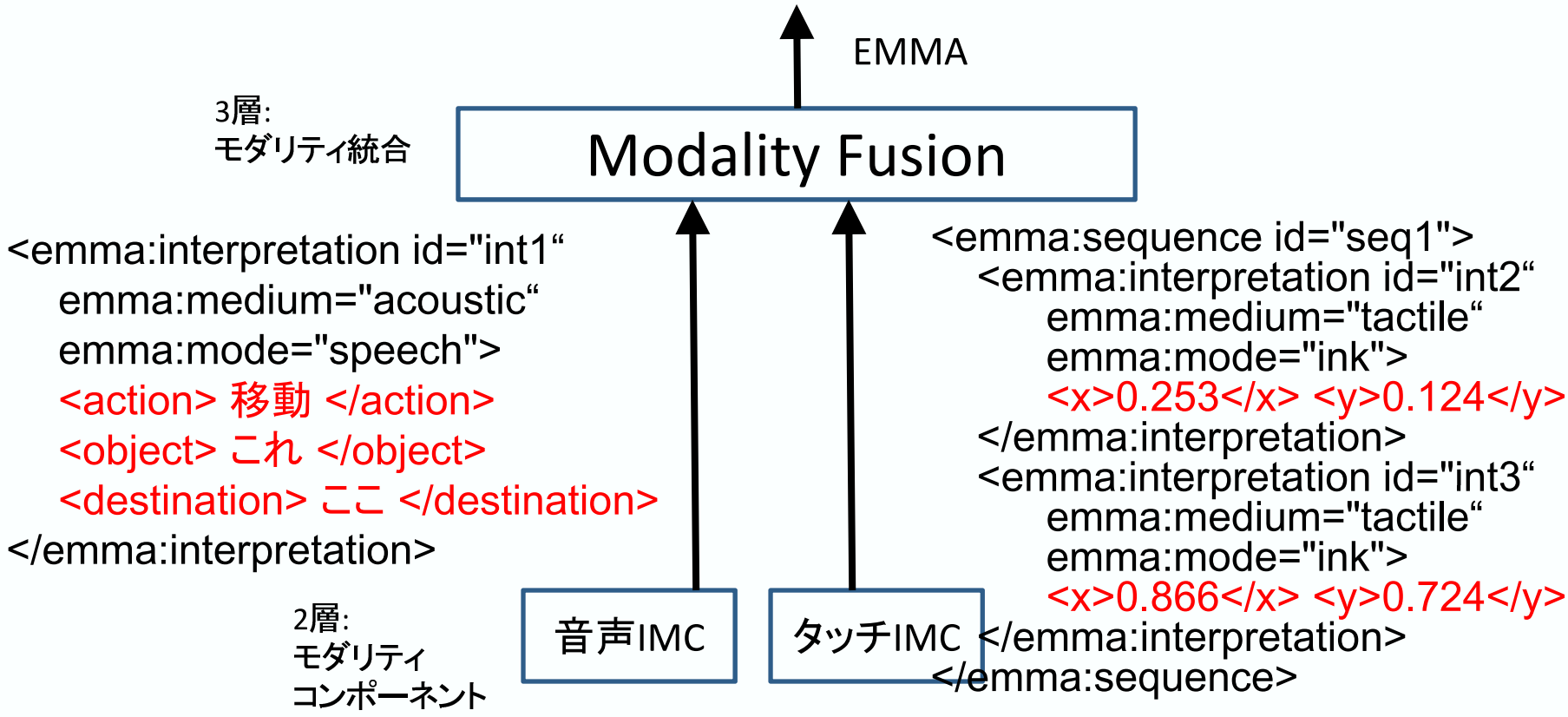
- 音声認識の即時結果表示の実現



3層：モダリティ統合

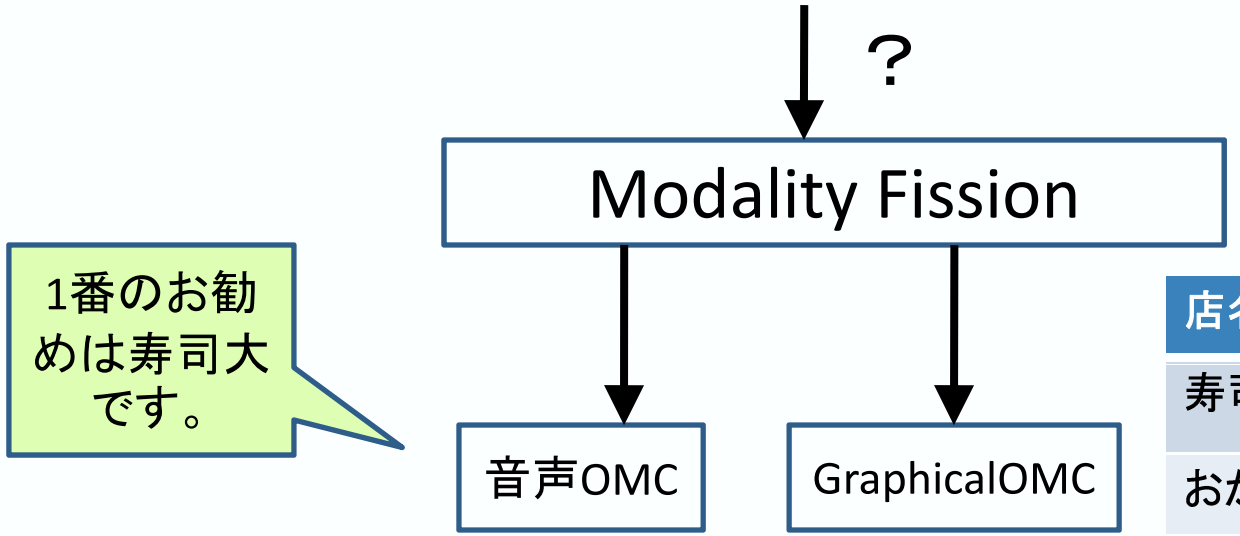
- 入力統合
  - 逐次入力や同時入力の解釈

例）「これをここに移動」+ペンタッチ2箇所



3層：モダリティ統合

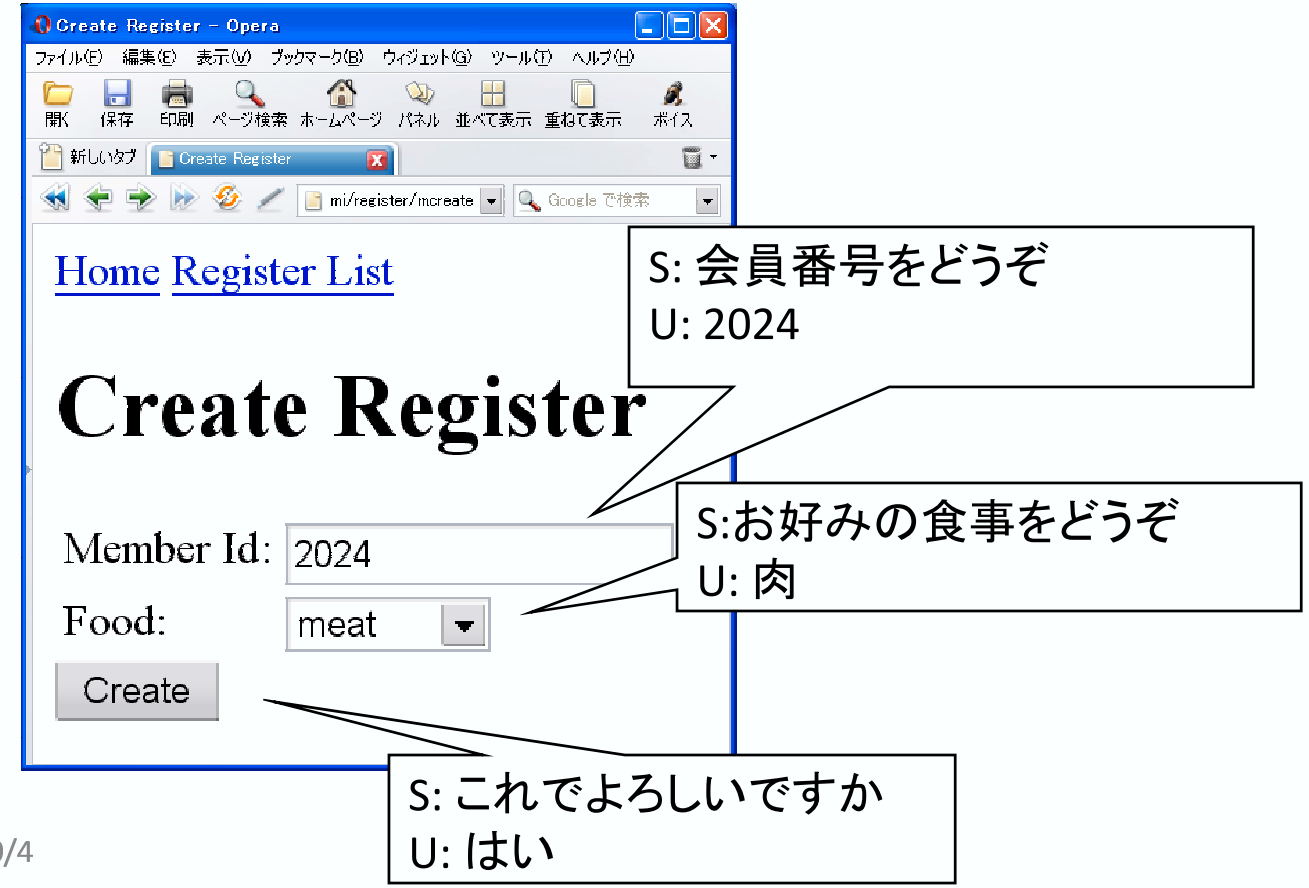
- 出力分化
  - 逐次出力や同時出力の同期
  - 利用可能なモダリティに応じて出力内容を調整
  - 何を入力とするかが研究課題



店名	価格	特徴
寿司大	3800円	きめ細やかな仕事
おかめ	3650円	2貫サービス
岩佐	3500円	貝尽くし

4層:タスク内制御

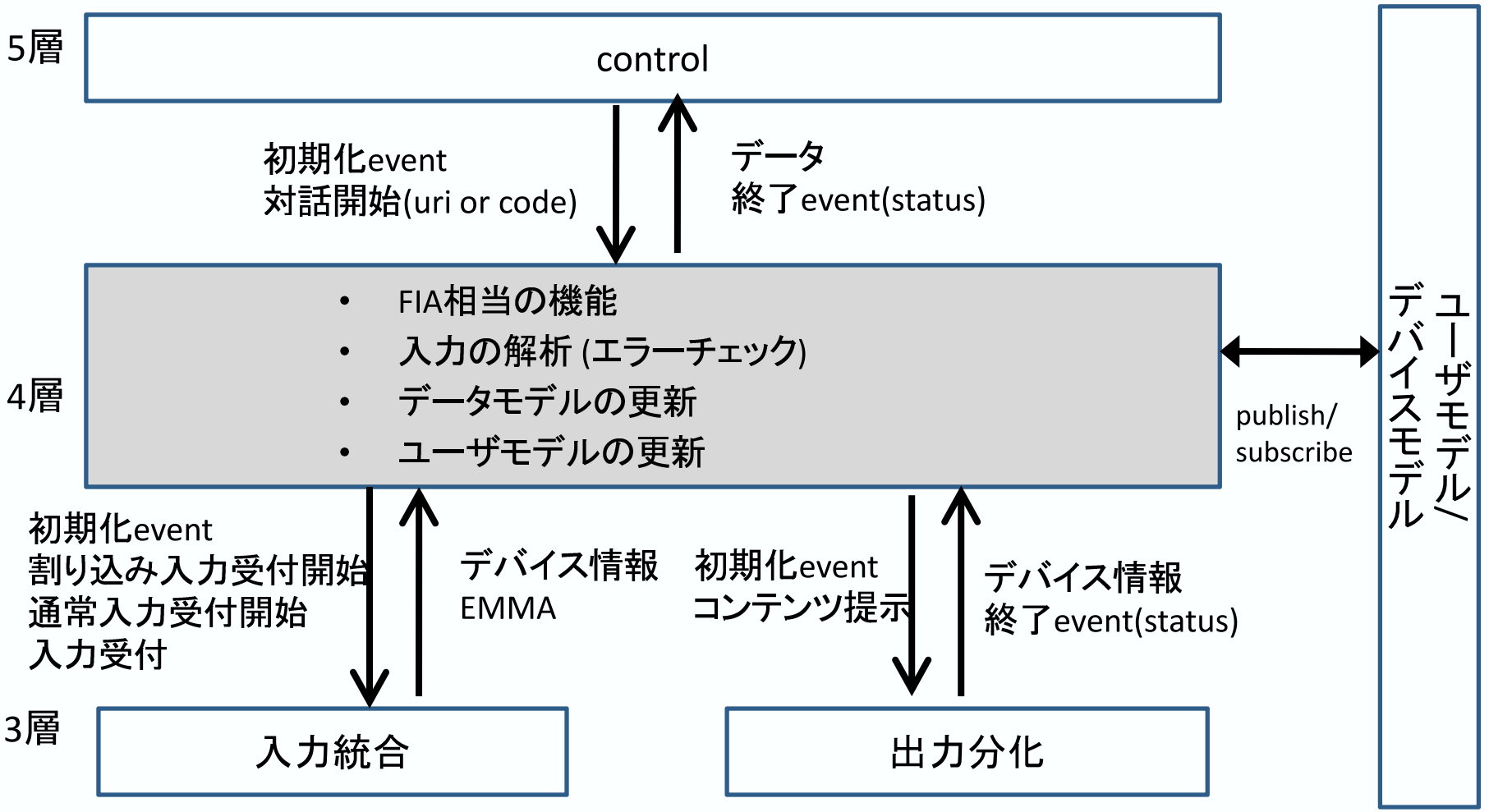
- イメージ
  - ひとまとまりの小さな対話タスク
    - クライアントサイドでの処理



4層：タスク内制御

- 必要な機能
  - エラーハンドリング
    - 例) 出発時刻＜到着時刻のチェック
  - 確認や再試行などのデフォルトメカニズム
  - フォームの充足性判定機能
    - 例）VoiceXMLのForm Interpretation Algorithm
  - スロット値の更新情報
    - 例） 確認に対する「いいえ、〇〇です」の処理

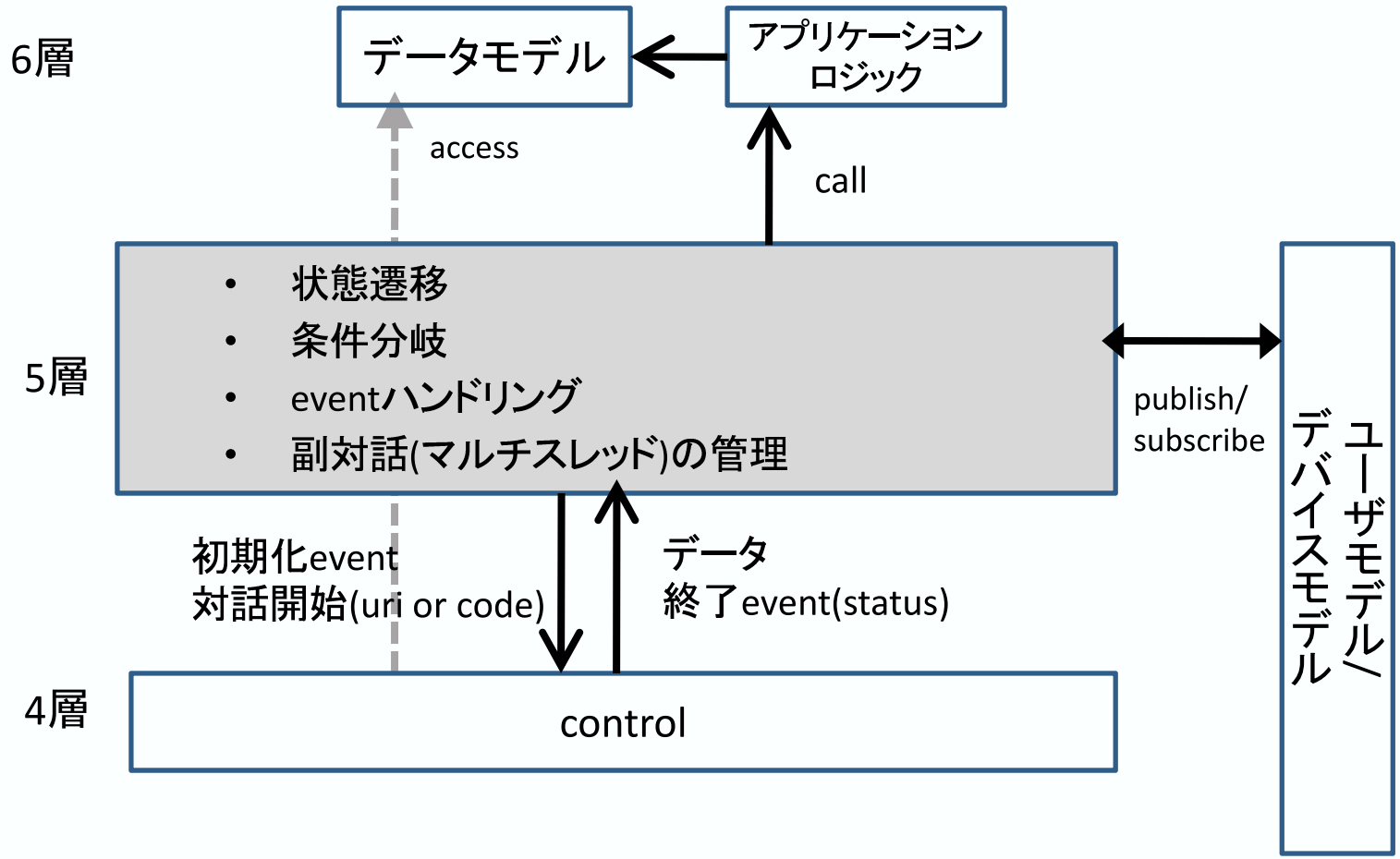
4層：タスク内制御


5層
control
初期化event
対話開始(uri or code)
データ
終了event(status)
4層
• FIA相当の機能
• 入力の解析 (エラーチェック)
• データモデルの更新
• ユーザモデルの更新
publish/
subscribe
ユーザモデル／
デバイスモデル
初期化event
割り込み入力受付開始
通常入力受付開始
入力受付
デバイス情報
EMMA
初期化event
コンテンツ提示
デバイス情報
終了event(status)
3層
入力統合
出力分化

5層:タスク間制御

- イメージ
  - タスクの大きな流れを記述
  - アプリケーションにアクセスし、その結果によって動的に対話の流れを変更
- 記述言語候補
  - SCXML(明示的に対話遷移を書く場合)
  - MVCのコントローラ記述
    - エントリーポイントとその処理を書く
  - Grailsでのgroovyスクリプト
    - アプリケーションロジックを包含

5層：タスク間制御


6層
データモデル
アプリケーション
ロジック
access
call
5層
• 状態遷移
• 条件分岐
• eventハンドリング
• 副対話(マルチスレッド)の管理
publish/
subscribe
ユーザモデル／
デバイスモデル
初期化event
対話開始(uri or code)
データ
終了event(status)
4層
control

6層：アプリケーション

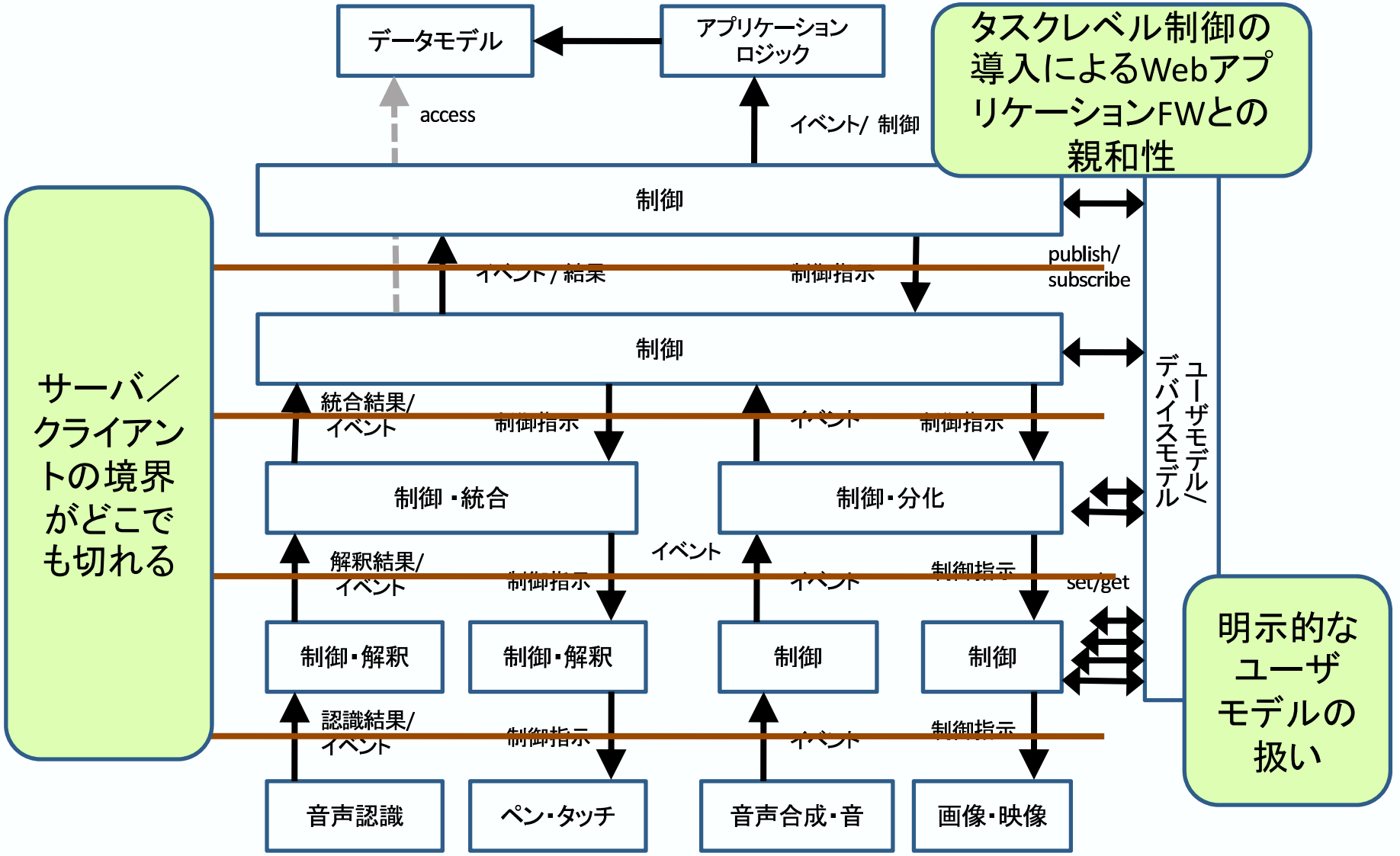
- 機能
  - 対話アプリケーションの外部のモジュール
  - アプリケーションロジック
    - 例）DBアクセス、Web APIアクセス
    - 情報の保存・更新・削除・検索　→Rails framework

ユーザモデル・デバイスモデル

- 共通の機能
  - セッションを越えて、対話アプリにユーザ情報・特性やデバイスの状態を通知
  - オントロジーで定義された変数の管理
- デバイスモデル

  cf.) W3C MMI 配信コンテクストコンポーネント
  - オントロジー+API
- ユーザモデル
  - デバイスモデルと同じ方法で実現したい

W3C MMIアーキテクチャとの違い



まとめ

- 情報規格調査会「音声入出力インタフェース委員会」の中間報告
  - MMIシステムのための階層モデルの提案
  - 各階層の仕様案
- 今後の予定
  - 「試行標準」案の公開へ
  - 参照実装・フレームワークの開発へ