

講演音声翻訳のための
多言語音声合成技術に関する研究開発
(091706004)

研究代表者

徳田恵一（名古屋工業大学）

研究分担者

李晃伸 南角吉彦（名古屋工業大学）

河井恒 倪晋富 志賀芳則（情報通信研究機構）

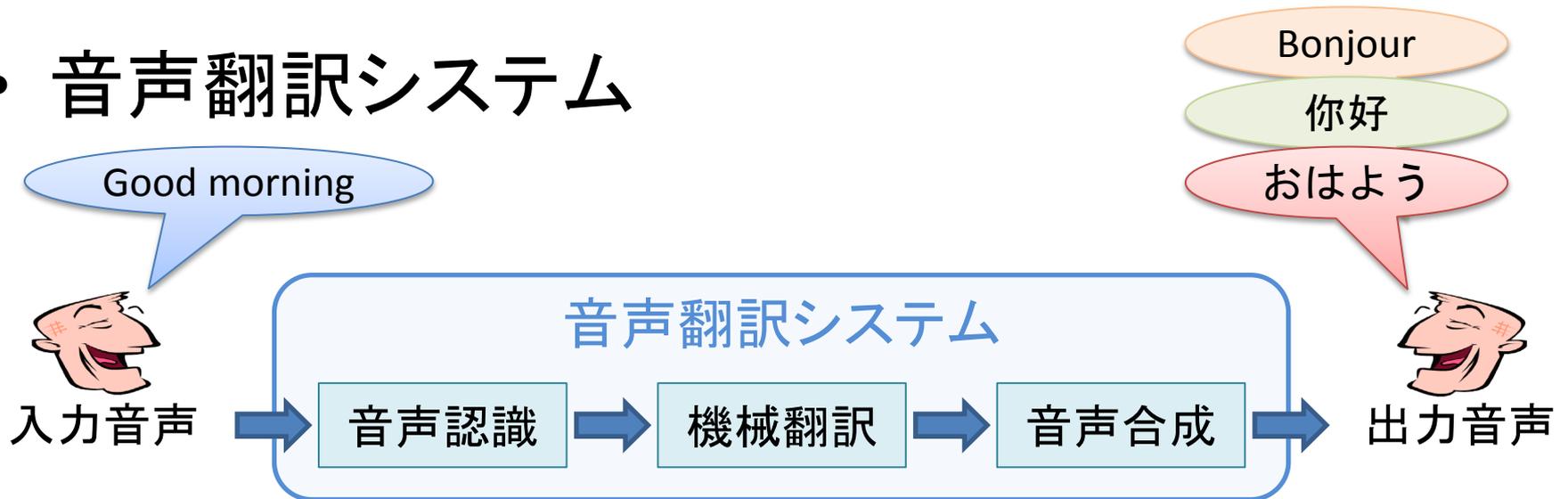
津崎実（京都市立芸術大学）

発表概要

1. 研究背景
2. 音声合成方式
3. 成果
4. その他の活動
5. まとめ

研究背景

- 音声翻訳システム



予め定められた話者の声で出力

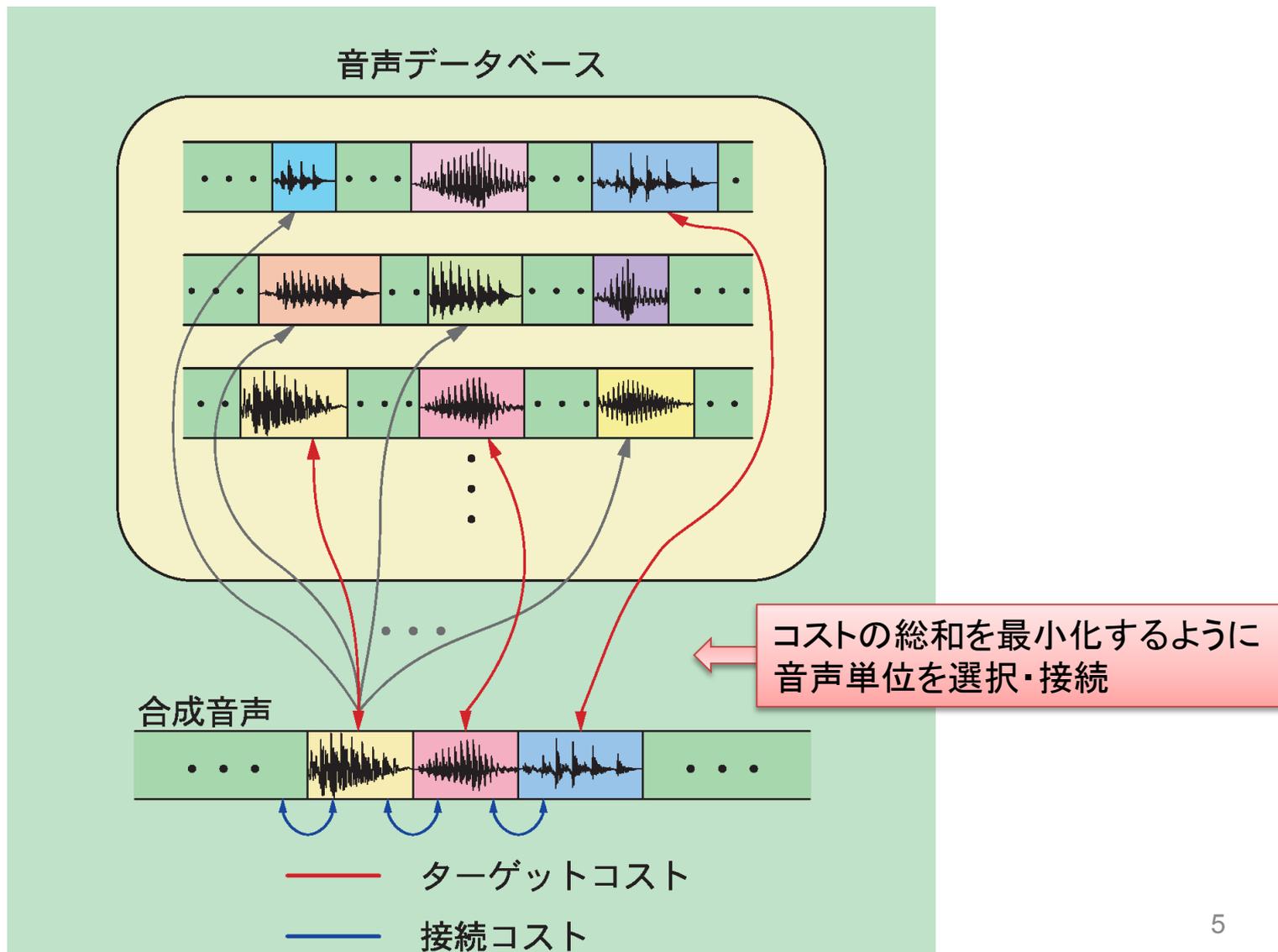
⇒ 円滑なコミュニケーションのために解決すべき問題

「入力音声の話者の声で、様々な言語の
音声を合成する」技術の研究開発

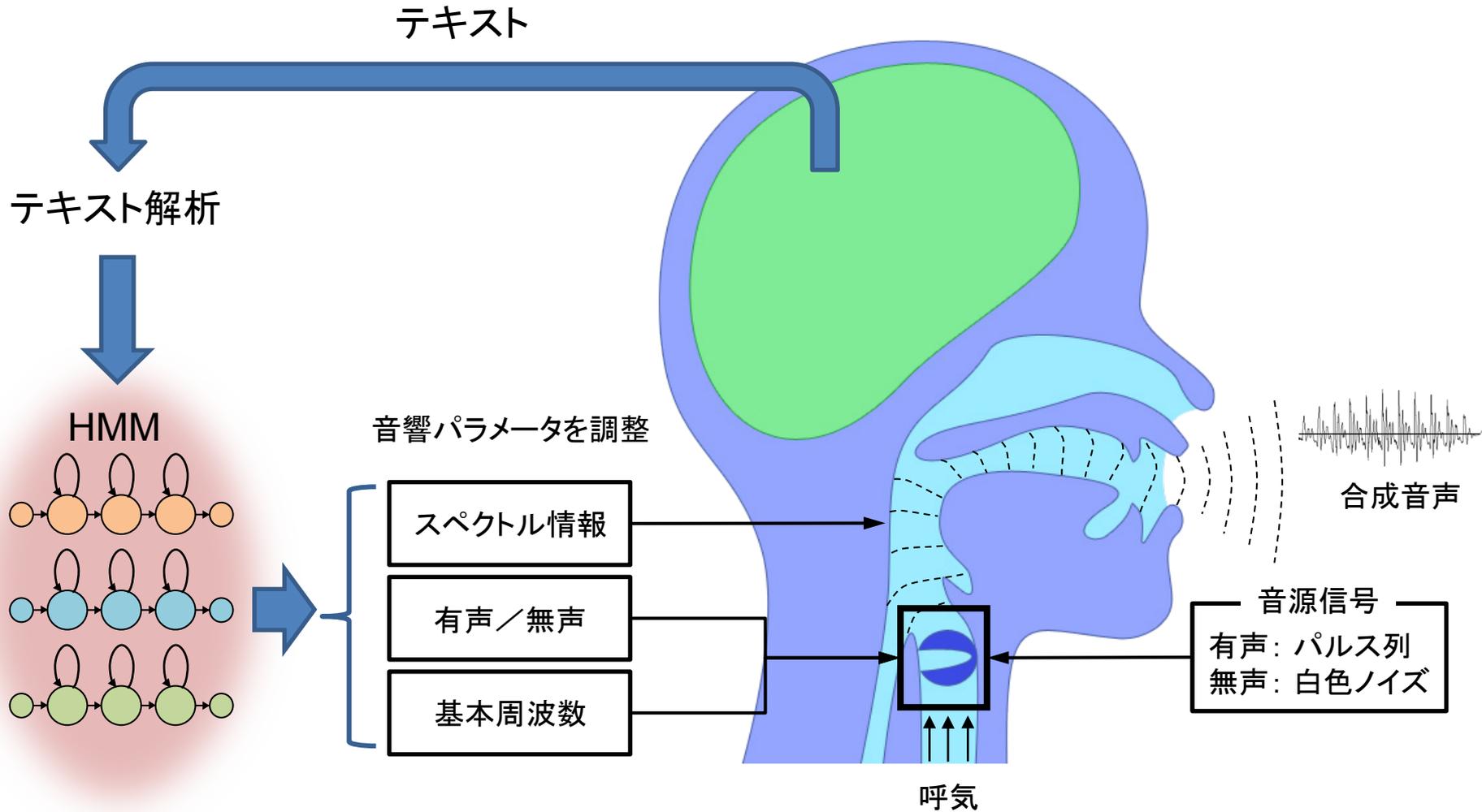
現在主流の音声合成方式

- 波形接続による音声合成
- HMM音声合成

波形接続による音声合成(従来法)



HMM音声合成



HMM音声合成

- HMM音声合成方式の特長

1. 統計モデル⇒多様な話者の音声を容易に実現可能
2. 低い言語依存性⇒多言語化が容易
3. 省メモリ⇒携帯デバイスでも実現が容易



次世代の音声合成方式として注目を集める

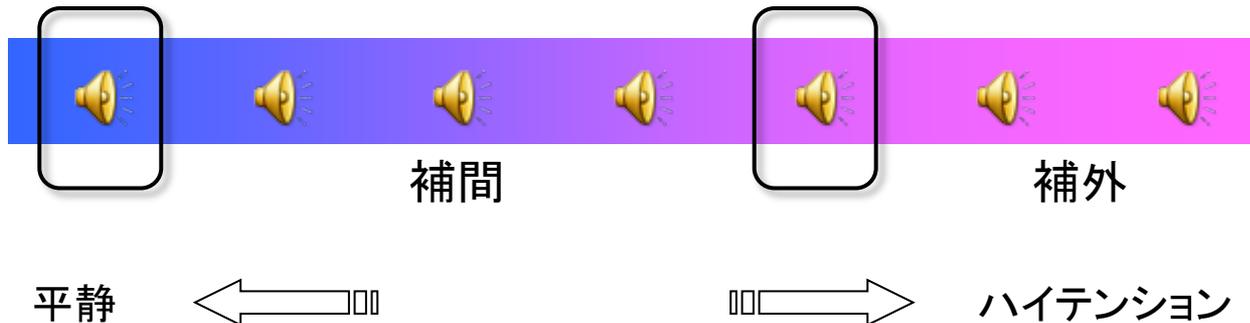
HMM: 隠れマルコフモデル (hidden Markov model)₇

HMM音声合成サンプル

- 読み上げ音声: 日本語 🗣️ 英語 🗣️ 中国語 🗣️
- 感情音声合成: 平静 🗣️ 怒り 🗣️
- 歌声合成: 🎵 🗣️
- 話者適応「声を真似る」: 
- 話者補間「声を混ぜる」:

「平静」のモデル

「ハイテンション」のモデル



成果

- 成果1:バイリンガル音声データベースの構築
- 成果2:言語間話者適応手法の確立
- 成果3:音声翻訳システムの試作
- 成果4:主観評価法の確立と評価
- 成果5:研究基盤ソフトウェアの整備

成果1:バイリンガル音声データベース の構築

- 日英・日中バイリンガル音声データベース

話者数	42
発話時間(発話文数)	42時間(27,189文)
発話内容	音素バランス文、合文法無意味文、 米国大統領演説原稿
発話スタイル	朗読、講演スタイル

※ 発声誤りの検査、発話の切り出し、ポーズ挿入位置の記録はすべて人手で行われた。

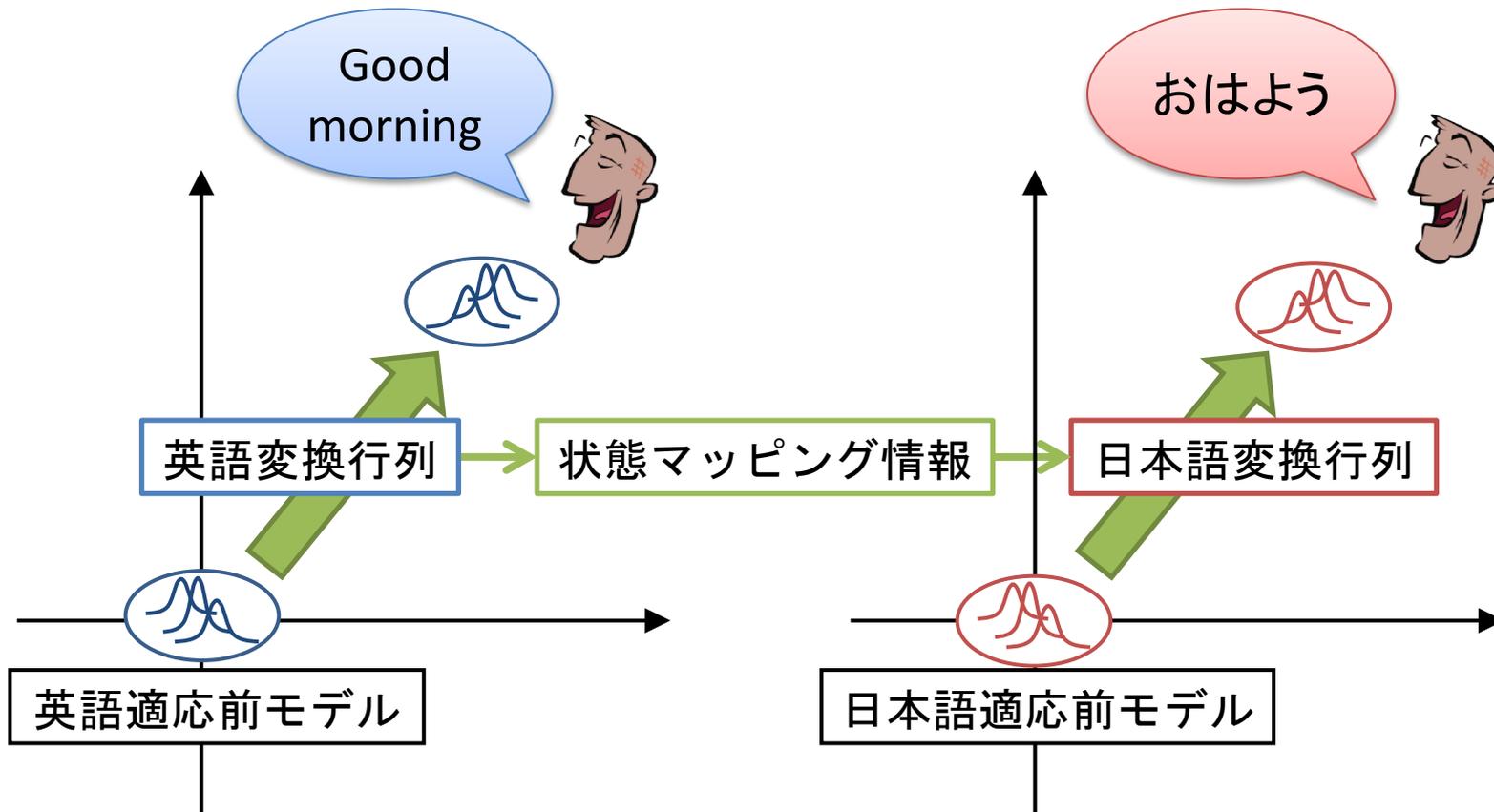
高度言語情報融合フォーラム(ALAGIN)の会員を対象として配布

成果2: 言語間話者適応手法の確立

1. 状態マッピングに基づく言語間話者適応
2. 固有声に基づく言語間話者適応

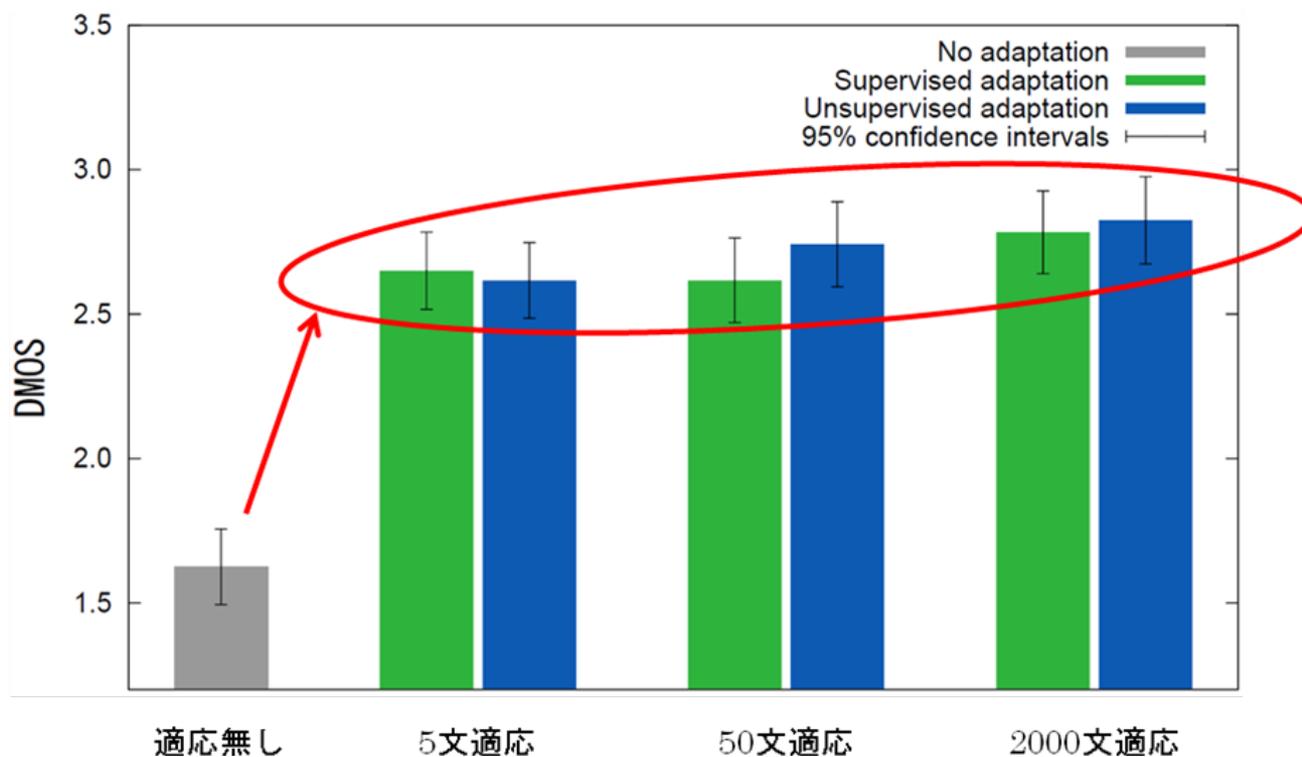
状態マッピングに基づく 言語間話者適応

- 言語間での統計モデル情報をマッピング



状態マッピングに基づく 言語間話者適応

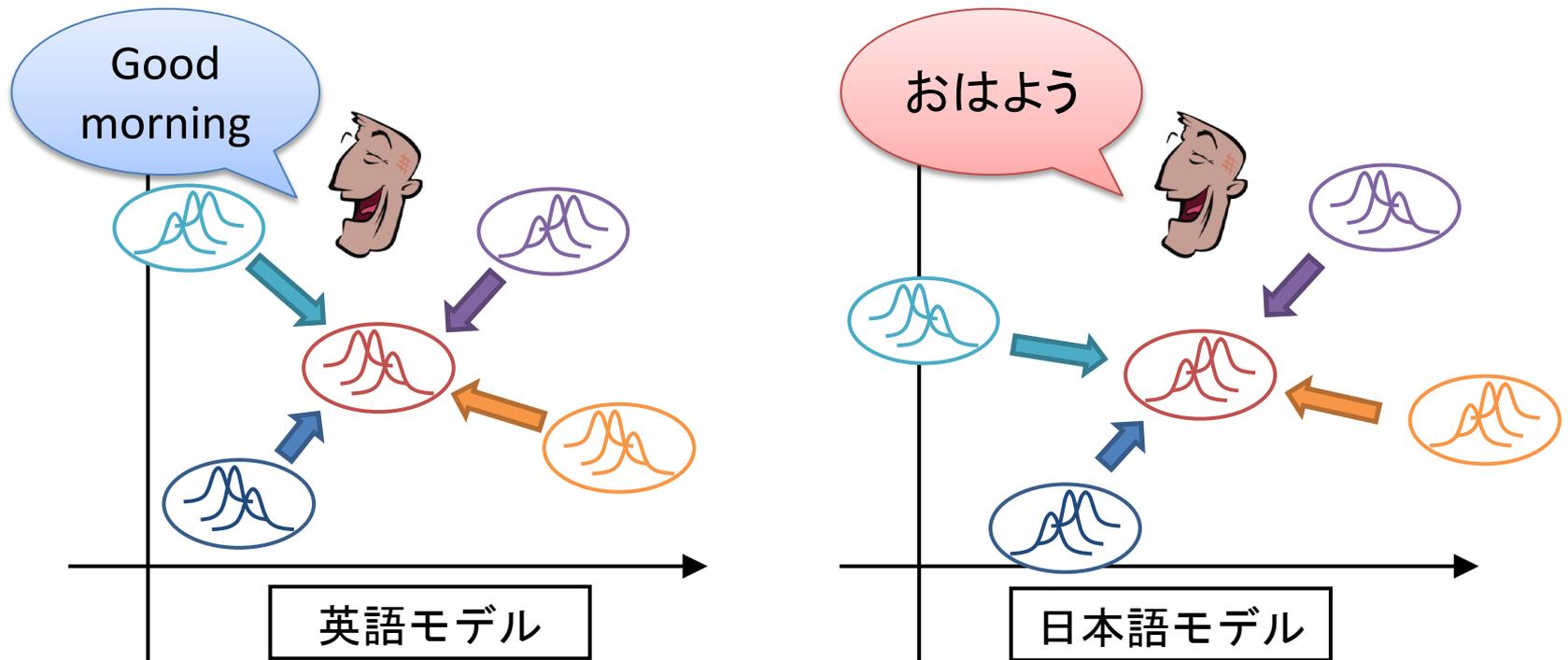
- 合成音声の話者類似性の主観評価



話者の声質の再現性を改善

固有声に基づく言語間話者適応

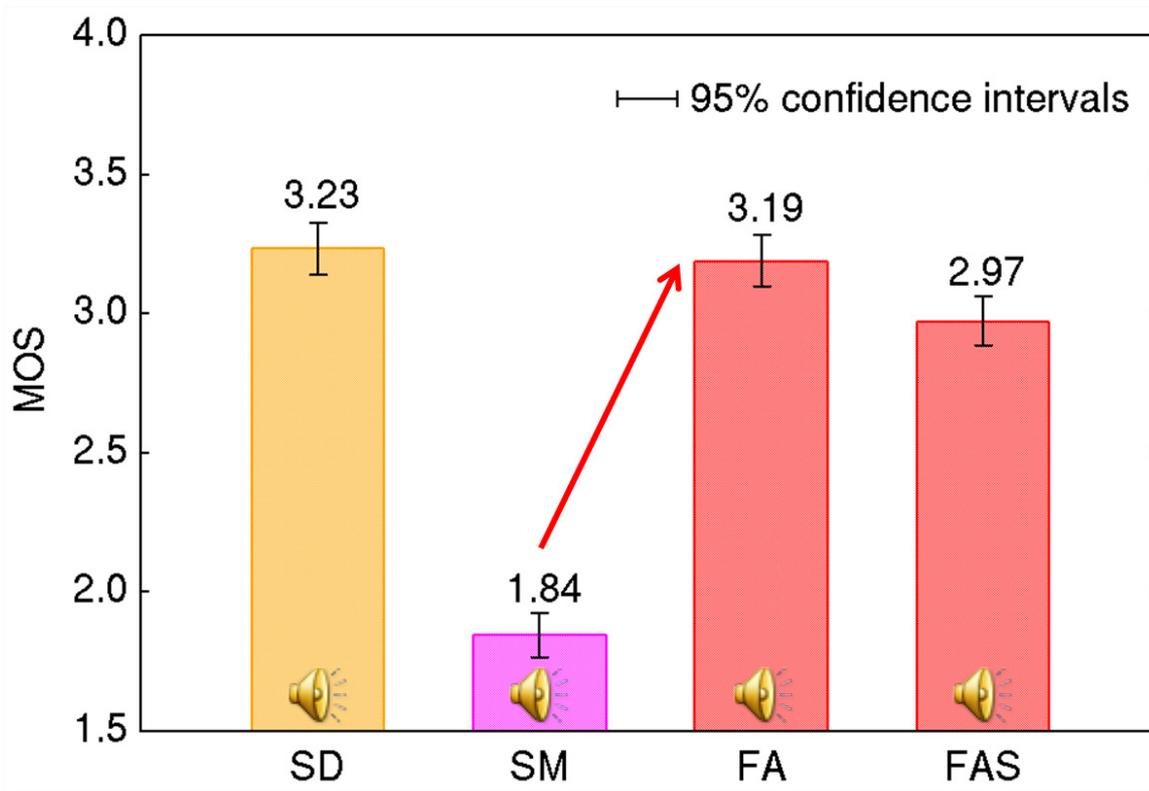
- 言語間で話者情報を共有



言語性と話者性を分離

固有声に基づく言語間話者適応

- 合成音声の自然性の主観評価

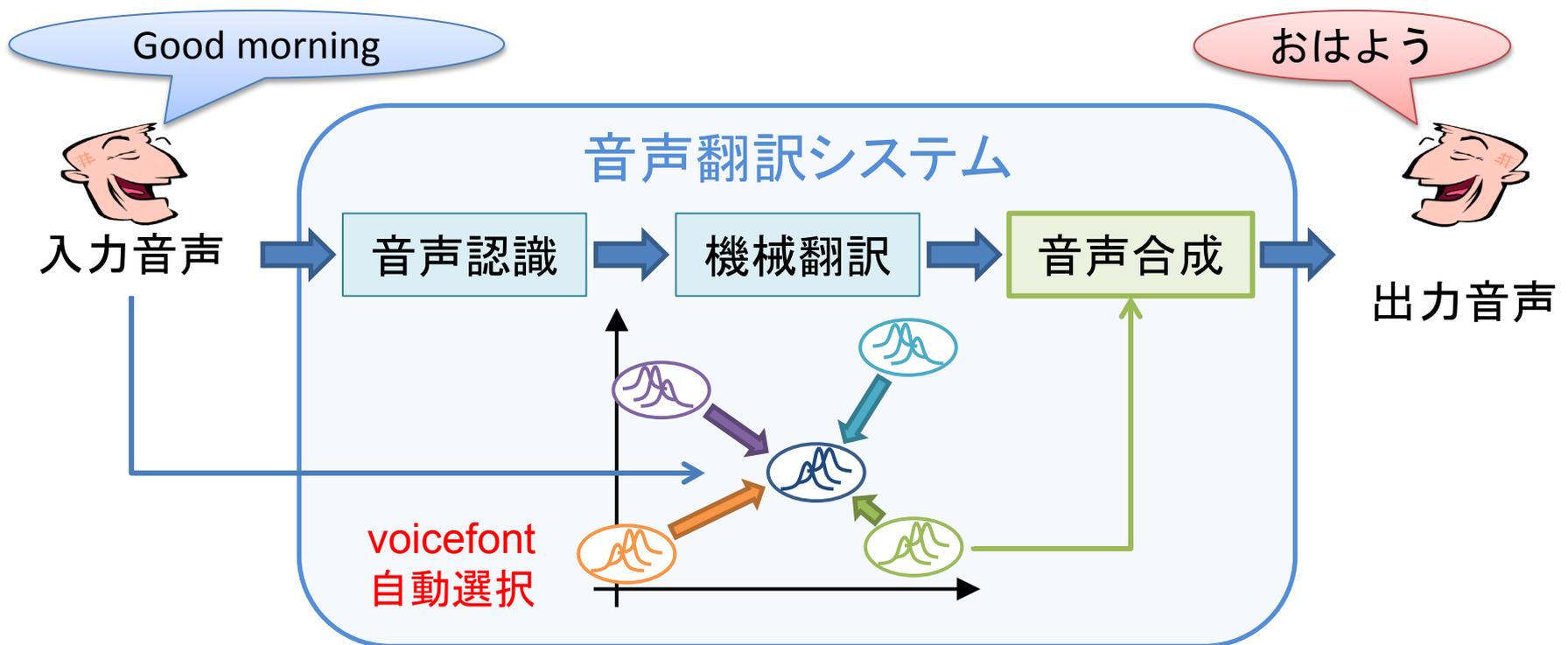


- SD: 特定話者モデル
- SM: 状態マッピングに基づく言語間話者適応
- FA: 固有声に基づく言語間話者適応
- FAS: 固有声に基づく言語間話者適応 (特徴量毎に因子推定)

自然性を大きく改善

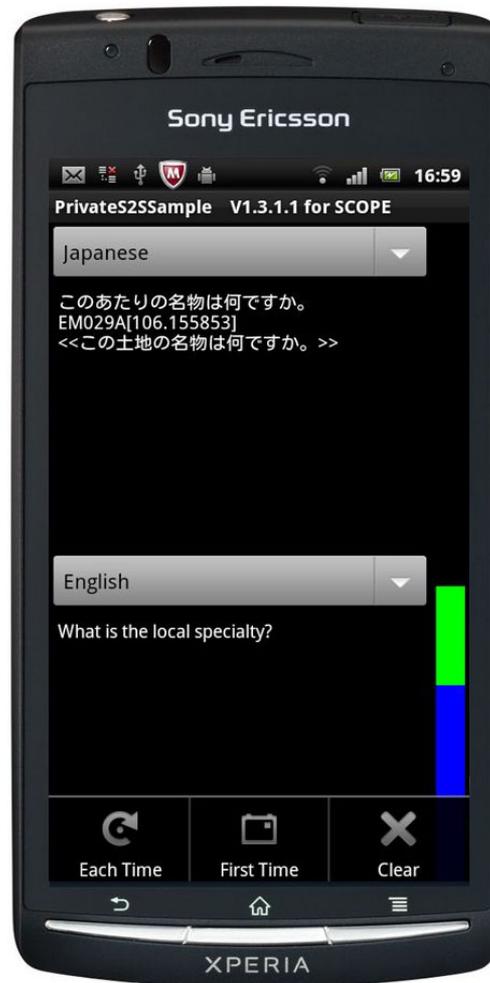
成果3: 音声翻訳システムの試作

- 音声合成用モデル(voicefont)自動選択機能
– 固有声に基づく言語間話者適応の簡易版



音声翻訳システムの試作

- デモビデオ
(1分半程度)



成果4：主観評価法の確立と評価

- 異言語間の話者性の判定
 - 通常の再認方略が使用不可能
 - 元話者に似ているかという印象を答えることは困難
 - ⇒ 2話者の自然音声と一方の合成音声を聞かせて
いずれを模擬したと思うかを選択させる
- 主観評価実験
 - 3種類の話者モデル選択法について比較
 - いずれの手法もチャンス・レベル以上の正答率
 - 話者の伝達性を改善

成果5：研究基盤ソフトウェアの整備

ソフトウェア	ダウンロード数 (直近1年間)	ページビュー (直近1年間)
大語彙連続音声認識エンジン 「Julius」	約9000回	約371000回
音声信号処理ツールキット 「SPTK」	約2000回	未観測
HMM音声合成ツールキット 「HTS」	約3000回	約60000回
HMM音声合成エンジン 「hts_engine API」	約1500回	約11000回
日本語音声合成システム 「Open JTalk」	約1400回	約8000回

※ hts_engine APIはNTTドコモの携帯電話20機種以上に採用

情報通信産業界へ大きく貢献

その他の活動

- 研究成果の普及に関する活動

査読付き論文	59件
口頭発表	72件
受賞	7件(+1件)
報道発表	6件
ワークショップの主催	5件

(+1件: 文部科学大臣表彰科学技術賞(研究部門)2012年4月17日)

- 人材育成

学部生	25名
修士課程学生	30名
博士課程学生	7名

まとめ

- 講演や演説等のモノローグを対象とした音声翻訳の実現に向けて「**入力音声の話者の声で、様々な言語の音声を合成する**」ための技術の研究開発
 - 成果1:バイリンガル音声データベースの構築
 - 成果2:言語間話者適応手法の確立
 - 成果3:音声翻訳システムの試作
 - 成果4:主観評価法の確立と評価
 - 成果5:研究基盤ソフトウェアの整備