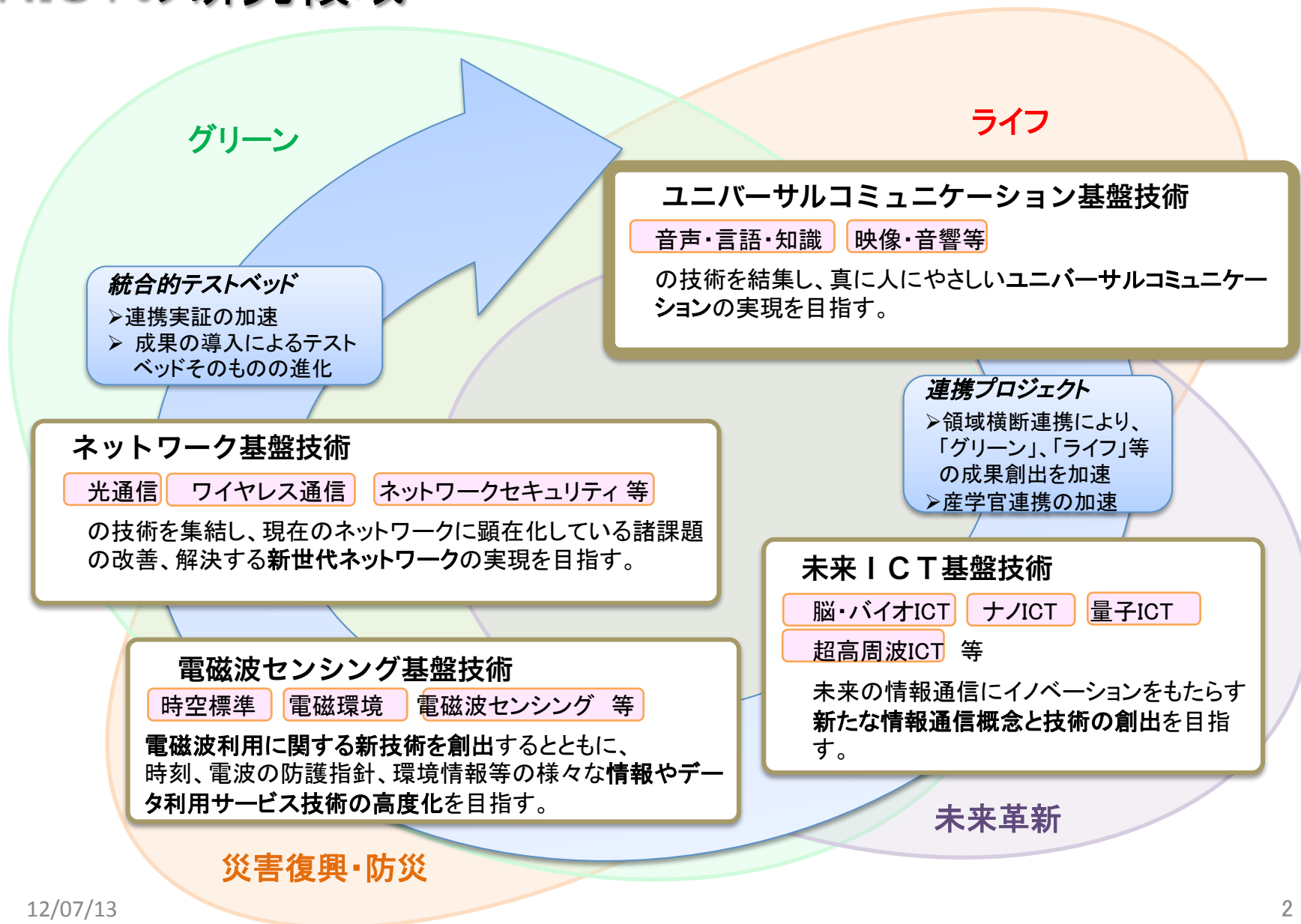


ネットワーク型音声翻訳(VoiceTra)及び 音声対話による観光案内(AssisTra)等 スマートフォン用アプリケーションの 研究開発

独立行政法人 情報通信研究機構
ユニバーサルコミュニケーション研究所
音声コミュニケーション研究室 室長

柏岡 秀紀

NICTの研究領域

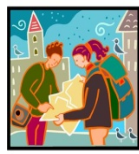


ユニバーサルコミュニケーション研究所における研究開発と社会還元

音声翻訳



高精度な逐次翻訳



多言語
音声対話

観光
通訳

国際産業
情報分析

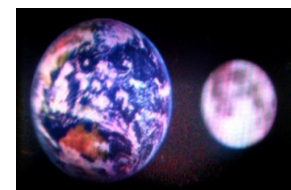
多言語
コミュニケー
ション技術

超臨場感
コミュニケー
ション技術

ユニバーサルコミュニ
ケーション基盤技術

コンテンツ・サービス
基盤技術

超臨場感メディア



200インチ裸眼 電子ホログラフィ
立体ディスプレイ

多感覚
インタラクション



地球環境
観測データ



Web

映像・音声
コンテンツ



収集

多言語翻訳技術

情報分析技術

大規模情報管理技術

管理・分析・
翻訳等

高度情報資産

サービス化

知識・言語
グリッド

提供

評価
利用

URCF

ALAGIN

けいはんなオー
プンラボ

社会還元

コンテンツ・サービス利活用

情報分析システム

自然な音声で簡単に情報を取得できるシステムを目指して 公開中の成果

NICT提供

- **VoiceTra**

旅行会話音声翻訳アプリ



- **ChaTra**

多人数対応型旅行会話音声翻訳アプリ



- **AssisTra**

京都観光案内音声対話アプリ(日本語版)



- **Kyo-no Hanna**

京都観光案内音声対話アプリ(英語版)



- **京のおすすめ**

京都観光スポットの紹介アプリ



- **WISDOM** (Web)

情報分析システム

<http://wisdom-nict.jp/>

NICTの技術移転による

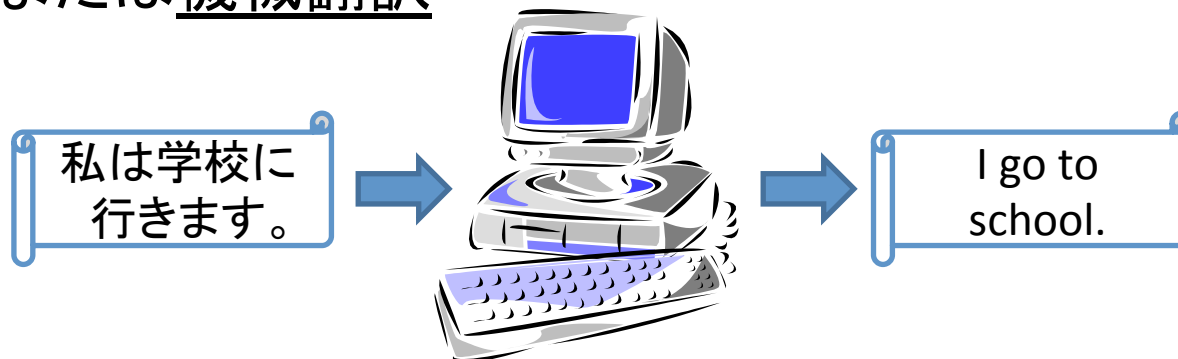
- **NariTra**

成田空港が提供する音声翻訳アプリ

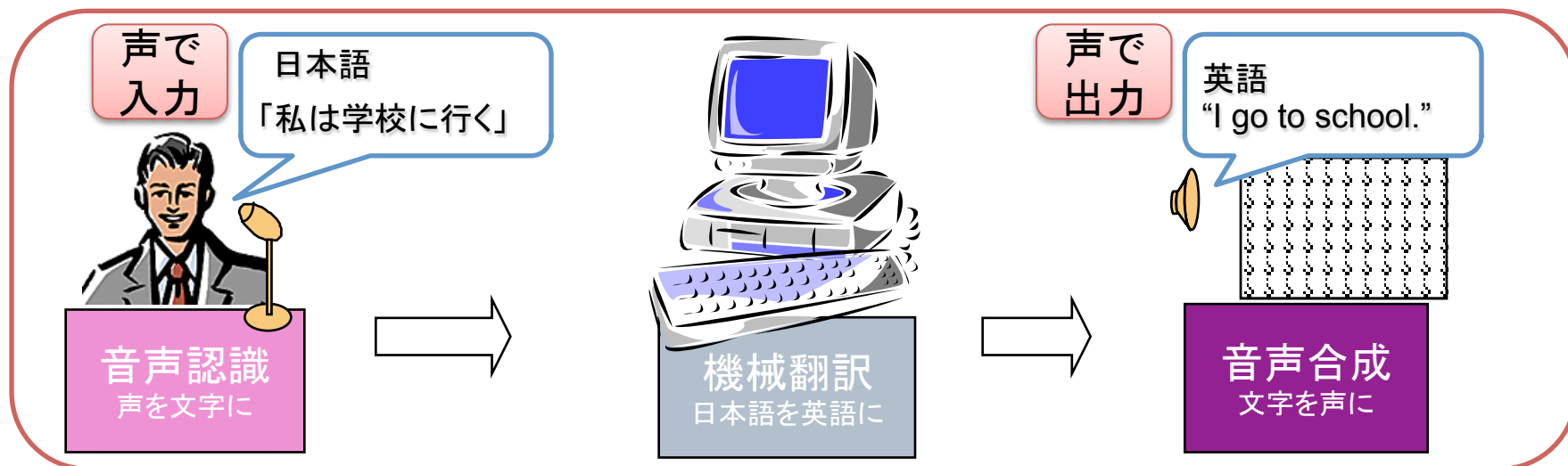


音声翻訳とは

自動翻訳または機械翻訳



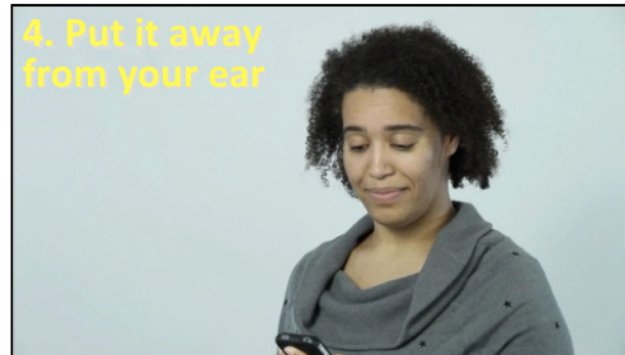
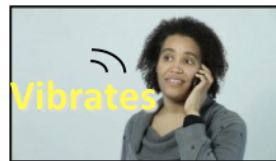
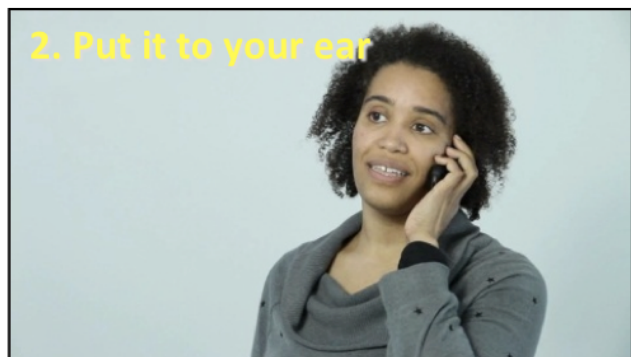
これに音声がつくと、



ネットワーク型(クラウド型)自動音声翻訳システム



VoiceTra

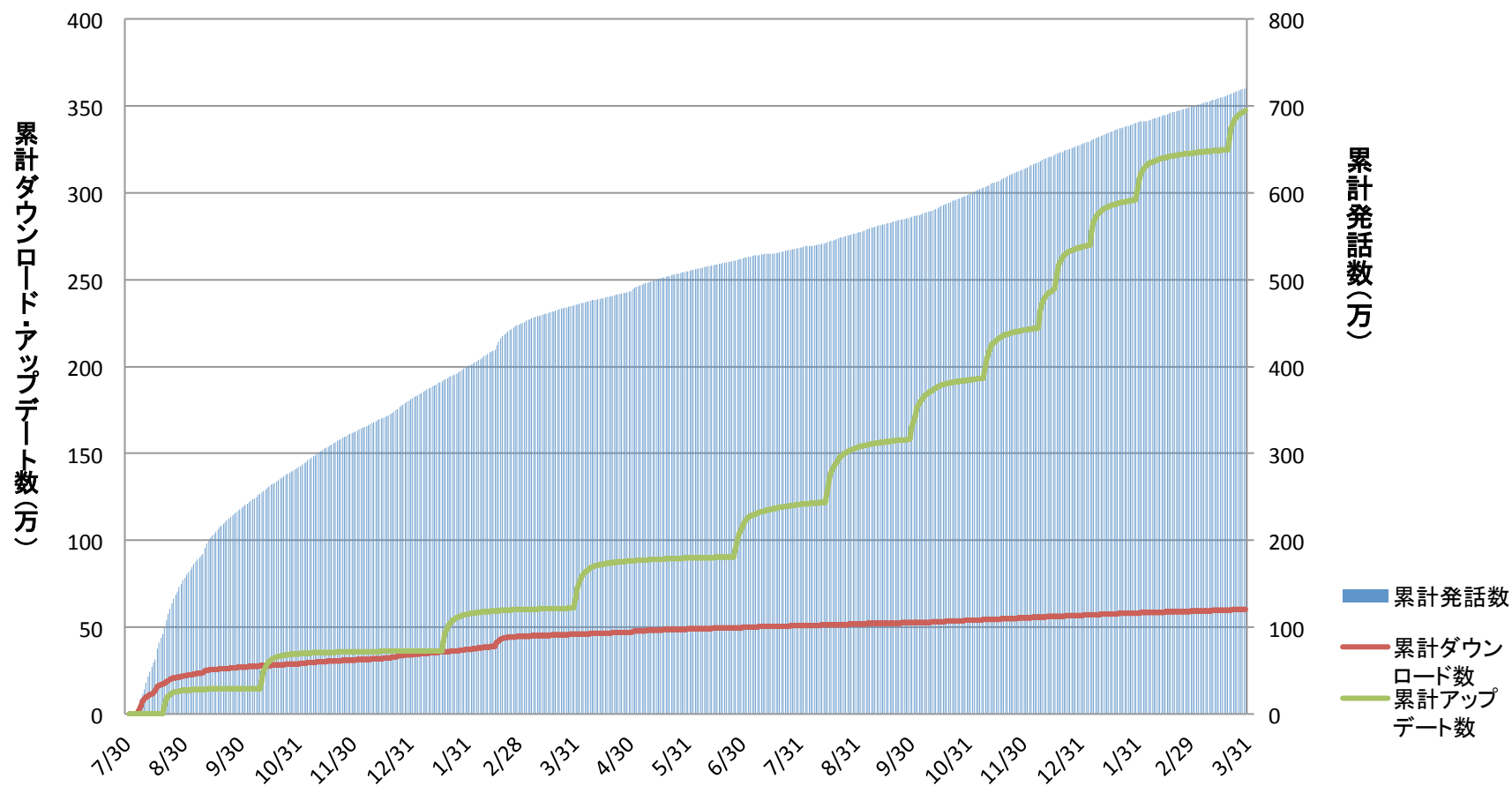


That's all



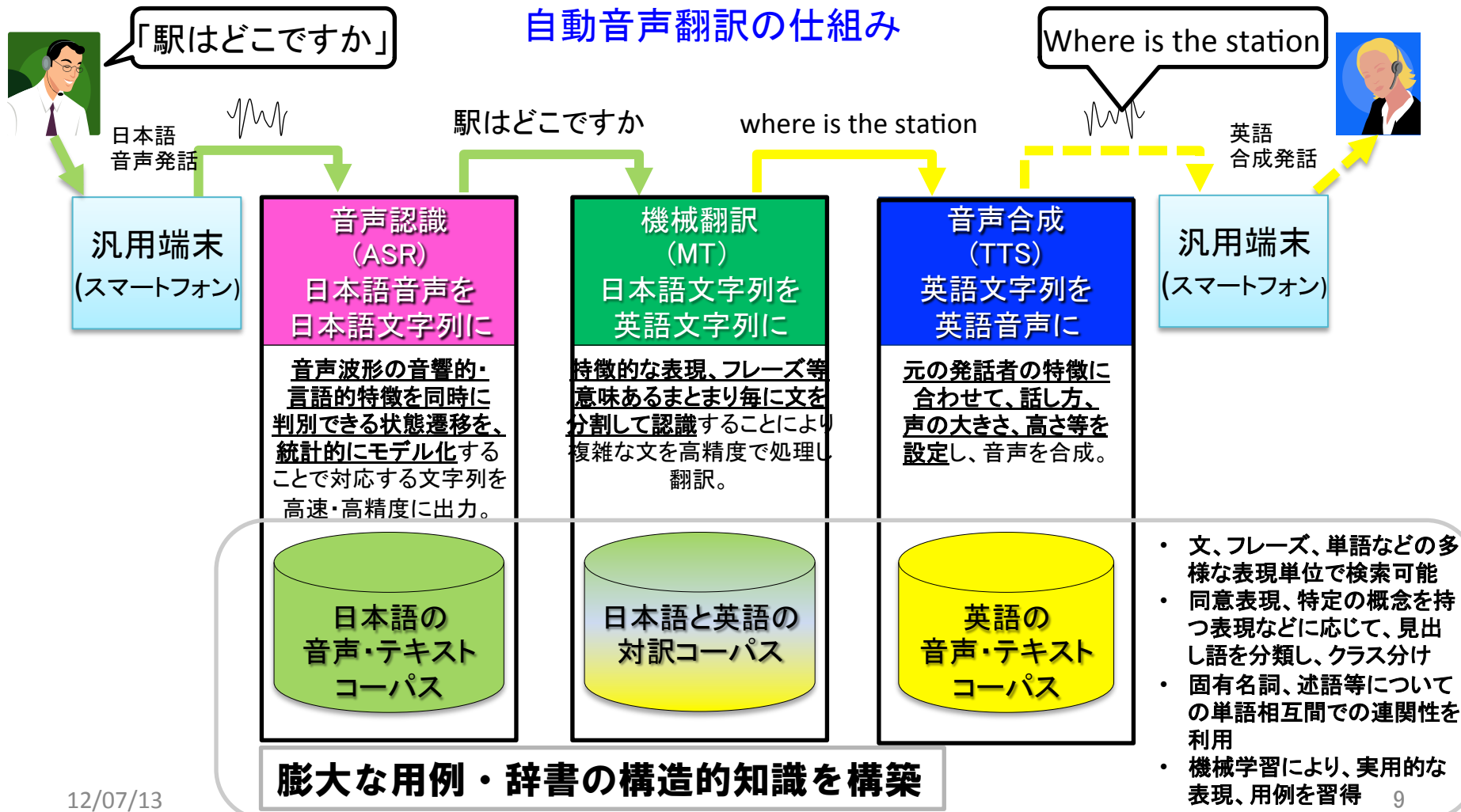
VoiceTraの実績

累計発話数とダウンロード、アップデート数



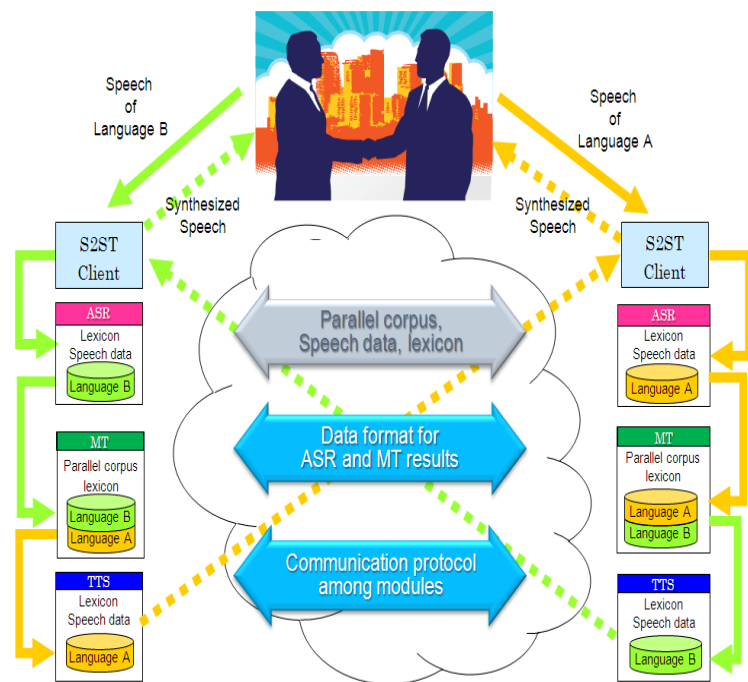
ネットワーク型多言語音声翻訳システム

- ・翻訳結果を機械学習することにより翻訳精度等が向上（「使うほど賢くなる。」）
- ・クラウドを利用し、汎用端末（スマートフォン）で利用可能
- ・ネットワーク上に分散する知識を活用し、幅広い話題に対応可能



ネットワーク型音声翻訳通信プロトコル標準化

NICTの研究成果をもとに、ITU-T※で国際標準化を実現(2010年10月)



ITU-T勧告F.745

音声翻訳サービスに必要な機能モジュール(音声認識、機械翻訳及び音声合成)をネットワーク接続するための要求条件、アーキテクチャ等を規定

ITU-T勧告H.625

機能モジュール間での通信を実現するためのインタフェース、プロトコル及びデータフォーマットを規定

これらの標準に準拠すれば、
ネットワーク上に分散した各機能を接続して
音声翻訳サービスを提供することが可能となる。

※ ITU-T: 国際電気通信連合 電気通信標準化部門

International Telecommunication Union - Telecommunication Standardization Sector

多言語音声翻訳システムのための国際研究コンソーシアム

多言語音声翻訳を一機関で構築することは非常に困難なため、世界各国の研究機関が参加する国際研究コンソーシアム「U-STAR」により、世界規模の音声翻訳研究ネットワークを実現

- ◆ メンバー各国において多言語音声翻訳のサーバーを構築
- ◆ スノーボール効果で、標準の普及、音声データの取得を促進

Universal Speech Translation
Advanced Research
Consortium: U-STAR

<http://www.ustar-consortium.com/index.html>



加盟機関: 21カ国の23機関(2012年4月現在)

NICT(日本)、ETRI(韓国)、NECTEC(タイ)、BPPT(インドネシア)、CASIA(中国)、CDAC(インド)、IOIT(ベトナム)、I2R(シンガポール)、DITT(ブータン)、KICS-UET(パキスタン)、LTK(ネパール)、MUST(モンゴル)、NUM(モンゴル)、UCSC(スリランカ)、UPD(フィリピン)、LIMSI(フランス)、PPCU(ハンガリー)、UULM(ドイツ)、シェフィールド大(イギリス)、PJIIT(ポーランド)、INESC ID(ポルトガル)、BME/TMIT(ハンガリー)、TUBITAK(トルコ)



成田空港で実用化 -NariTraの概要-

- アプリ名: NariTra(ナリトラ) Narita+Translatorより。
- 対象言語: 日本語⇄英語、中国語、韓国語
- 対応端末(OS): iPhone (iOS4.0～)およびAndroid端末(Android2.2～)
- リリース日時: 2011年12月22日
- 入手方法: App StoreまたはAndroid Marketよりダウンロード
- 利用料金: 無料(※通信料金は別途かかります)
- 利用場所、利用者に関する制限: なし
- 音声翻訳技術: すべてNICTが提供。

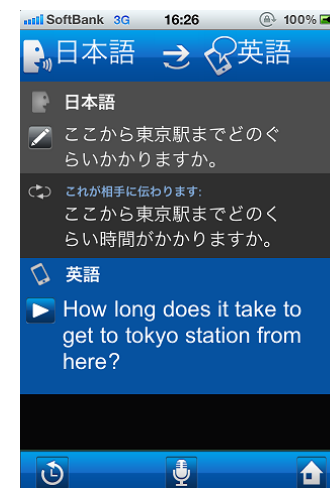
アプリアイコン



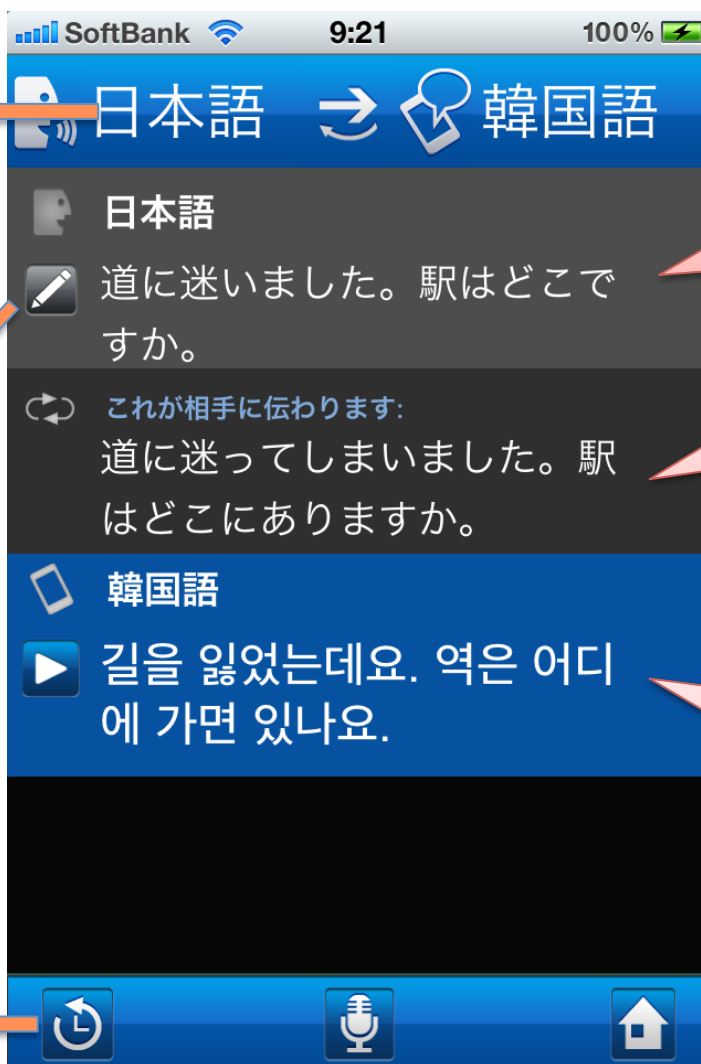
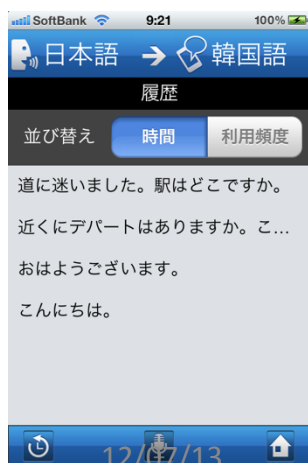
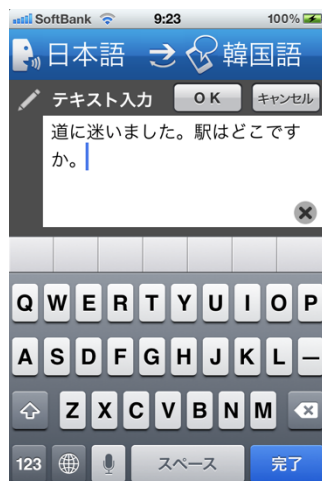
起動画面



翻訳画面



NariTra (主な機能)

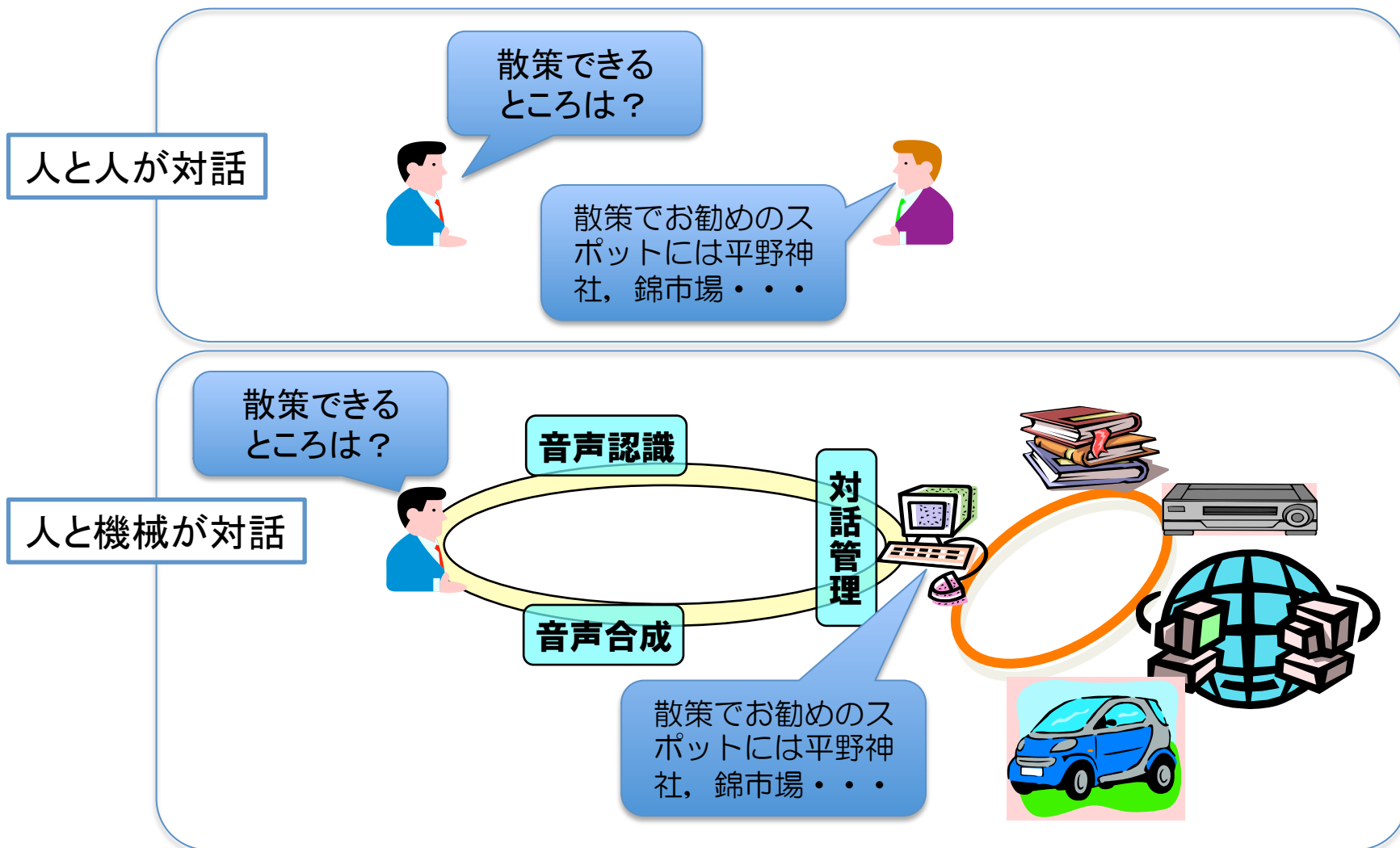


音声認識結果

逆翻訳結果
日→韓→日

翻訳結果

音声対話システムとは



京都観光支援ツール AssisTra(アシストラ)

NICTの音声対話処理技術による京都観光支援ツール

App Storeにて無料で提供 「AssisTra」で検索してみてください。

★三つのサービス:

- はんなのガイド 京都編
 - 「京野はんな※」が京都観光ガイドとして旅のお供をします。
あなたの声を理解し、京都の観光案内をします。
- KyoTra
 - 京都の観光名所の検索や近隣の観光地を調べることができます。
日英中韓の言語に対応しています。
- TraMemo
 - 写真の撮影や撮影した写真から自分だけの観光スポットを作成できます。



※NICTユニバーサルコミュニケーション研究所が「けいはんな(京阪奈)」地区にあることから名づけられました

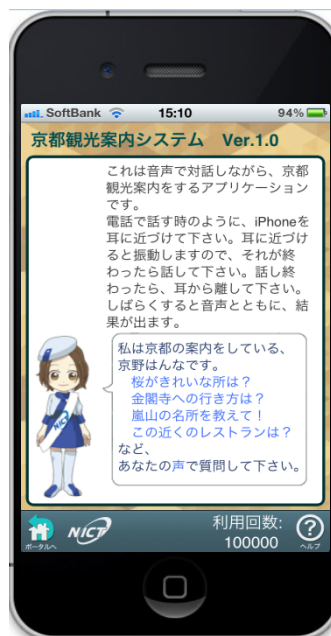


はんなのガイド 京都編

- 京都観光ガイドの「京野はんな※」が、音声によるユーザの要求を認識・理解し、観光スポットの推薦や情報の検索を行い、音声で応答します。

プロの観光ガイドと旅行者の模擬会話(約300対話)をモデル化し、対話制御機構を構築しています。

京都の主要観光スポット約100箇所についての様々な観光情報をご案内できます。現在は、日本語にのみ対応しています。



自然な表現で話しかけることで、必要な情報が得られますので、面倒な操作は必要はありません。

対話の自然性を実現するため、対話の履歴、位置情報等から、発話された状況に応じた理解、応答をします。

※NICTユニバーサルコミュニケーション研究所が「けいはんな(京阪奈)」地区にあることから名づけられました

音声で会話しながら気軽に観光情報検索

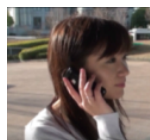


京野はんな

わたしが
ご案内します。

【使い方は簡単！】

- ①耳に近づけると震えます。
- ②電話するように話します。
- ③耳から離します。
- ④再度震えたら入力完了。
- ⑤しばらくすると、はんなが音声で応答します



散歩できる
ところは？

平野神社につ
いて教えて

地図を
見せて

行き方を教え
て



散歩でお勧め
のスポットに
は平野神社、
錦市場・・・

はい、平野神
社ですね。平
野神社は・・・

平野神社周
辺の地図を
表示します。

京都駅から平
野神社までの
行き方を表示
します。

主要観光スポット情報(桜や紅葉、
景色、庭園、仏像、建物などの情
報、拝観料、開館時間など)
周辺スポット検索
周辺レストラン検索
周辺地図検索
バス経路検索
などご案内できます。

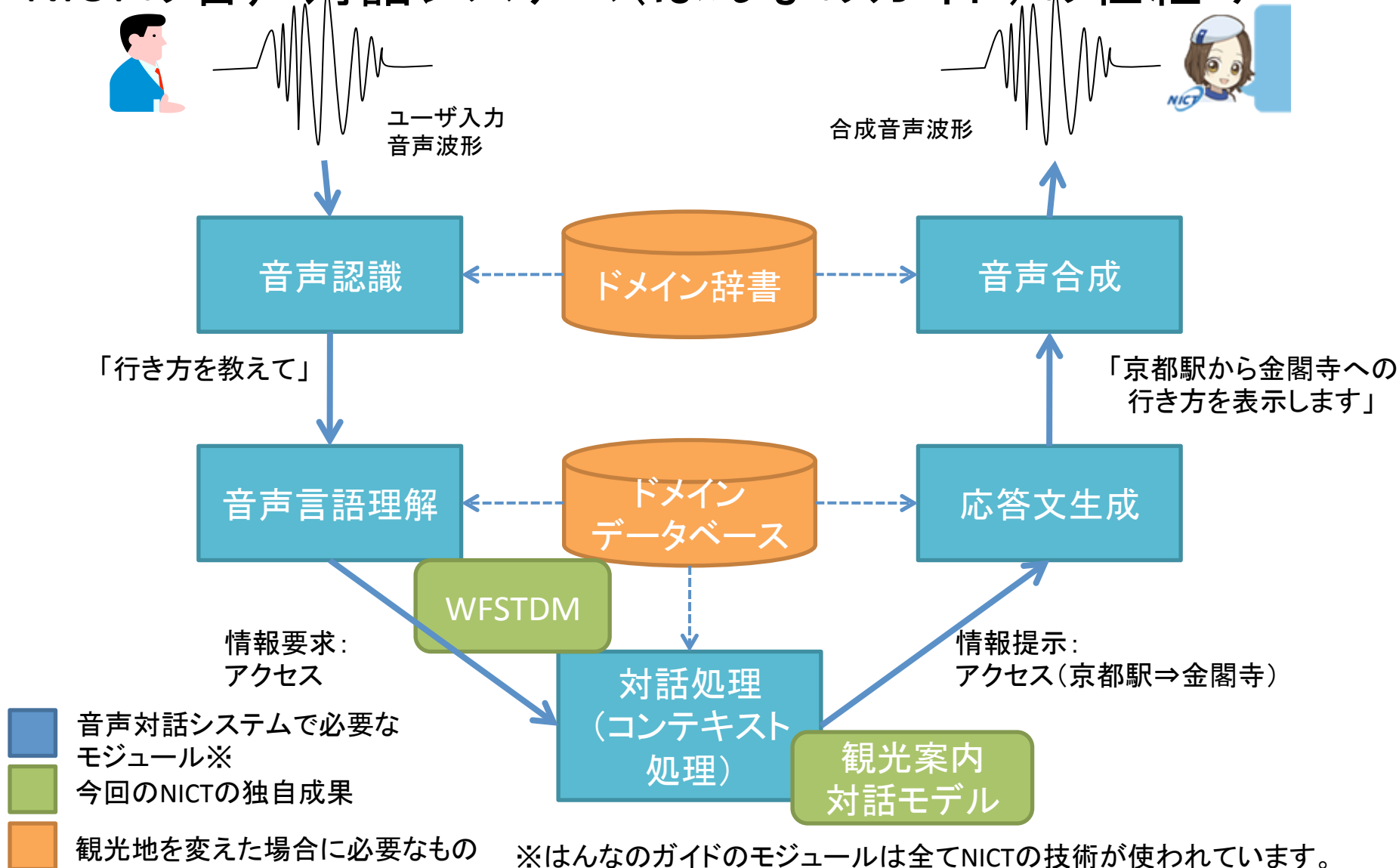


AssisTra (はんなのガイド)

smart phone用 音声対話アプリケーション



NICTの音声対話システム(はんなのガイド)の仕組み



インターネット上の情報などを活用し、文化的な背景や文脈を考慮した コミュニケーションを実現するナチュラル言語コミュニケーションの研究開発を進める

多言語で、多様な分野の情報を、多様な対象から、自然に取得するための基本技術開発

多様な分野で利用できれば

対象分野の拡張:

観光、医療、災害復興...

長い発話でも大丈夫かな？

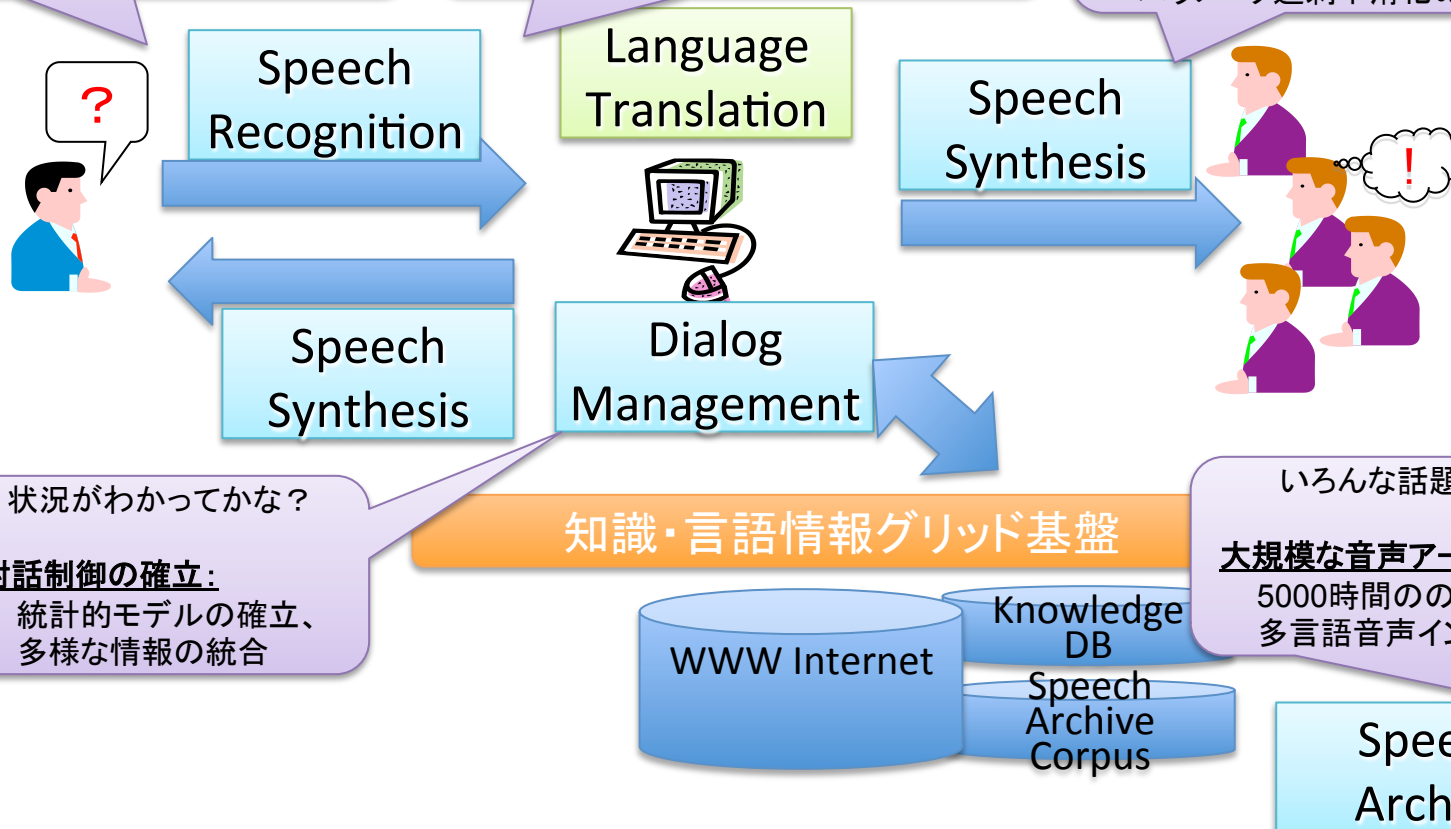
長文・高速音声認識:

言語モデル・認識アルゴリズムの改良

自然な発話をしてくれると！

高精度音声合成モデルの確立:

位相を考慮したモデル
パラメータ過剰平滑化の回避



状況がわかってかな？

対話制御の確立:

統計的モデルの確立、
多様な情報の統合

いろんな話題に対応してるかな？

大規模な音声アーカイブ構築:

5000時間のコーパス(現状の5倍)
多言語音声インデキシング技術開発

音声認識: SprinTra WFSTによる大語彙音声認識

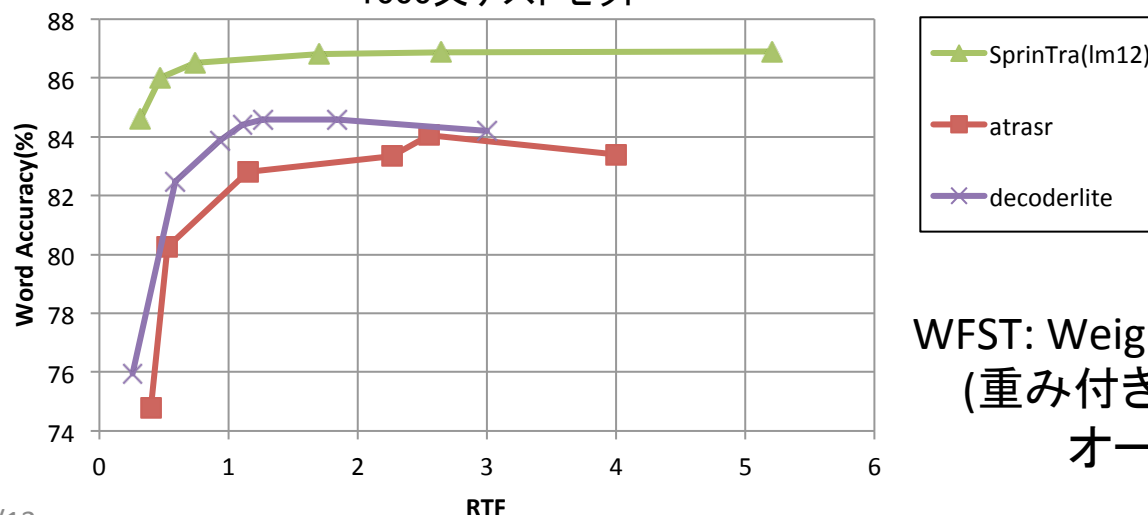
• 音声認識デコーダの開発

WFSTの最適化演算を用いることにより,

- 逐次的対応付けを解消・探索エンジン軽量化
- 複数モデルを統合・知識源全体を再整理

- 近年, 高速な音声認識を実現する手法として最も注目を集めているアプローチ
- 高速かつ高い汎用性と拡張性を持つ

SprinTraと既存のデコーダとの性能比較(32万語彙)
1000文テストセット



WFST: Weighted Finite-State Transducer
(重み付き有限状態トランスデューサ)
オートマトン(Automaton)の一種

音声認識性能を改善 (高齢者・子供音声)

子供から高齢者まで幅広い年齢層の話者に対応

◆ 音声コーパス収集

- 音響モデル学習用に、高齢者(60歳以上)、子供(小学校1年生～4年生)の音声を各100名以上収集

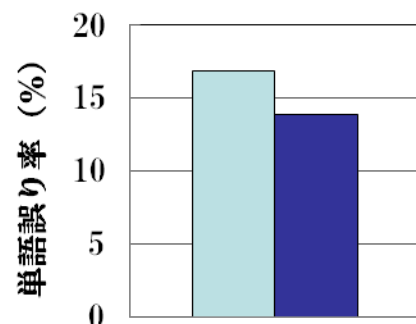
◆ 音響モデル構築

- 収集した音声コーパスをもとに、年齢層別音響モデルを構築
- 年齢区切り自動推定手法により、年齢層による変動を効率的にモデル化
- 複数モデルの並列探索により、話者の年齢が未知でも高精度に認識

◆ システムへの搭載

- 音声翻訳全国実証実験用のシステム(高齢者対応モデル)、日本科学未来館でのデモ展示用システム(子供対応モデル)に搭載

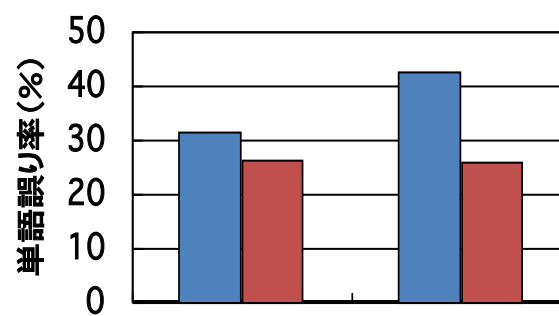
□ ベースラインモデル ■ 高齢者モデル



旅行会話音声

高齢者音声での評価

■ ベースラインモデル ■ 子供モデル



旅行会話音声 日本科学未来館

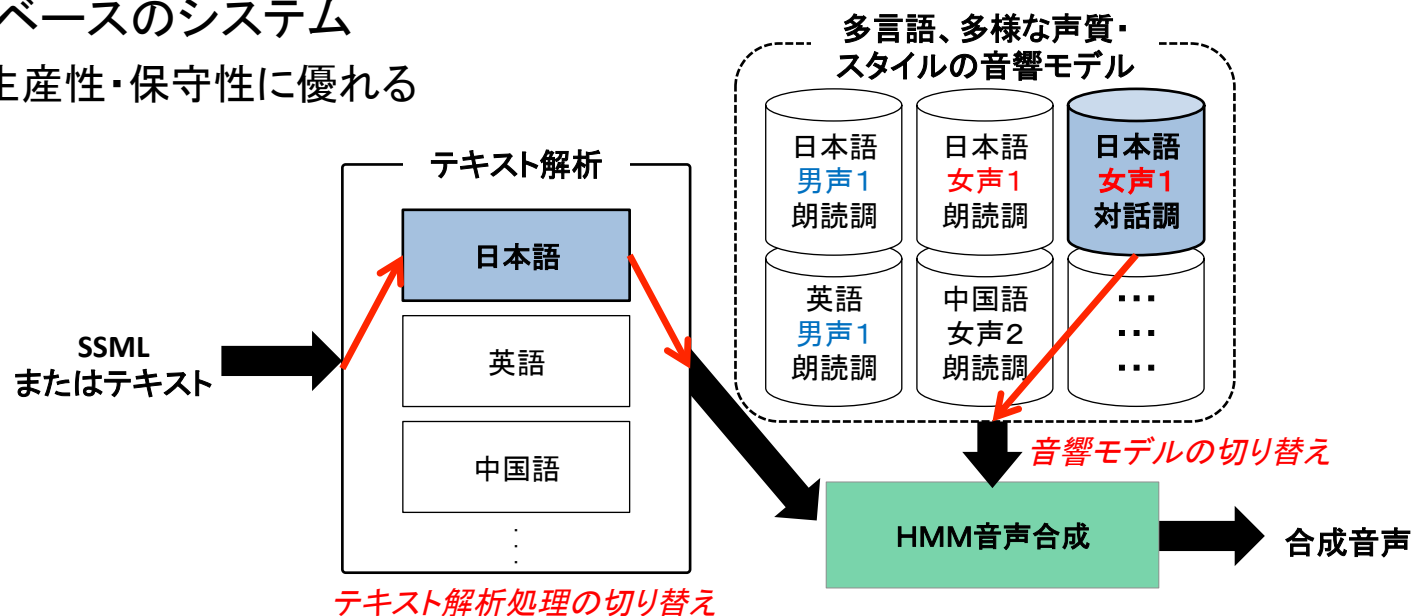
子供音声での評価



日本科学未来館で行われた
「ドラえもん科学未来展」
のでの利用風景

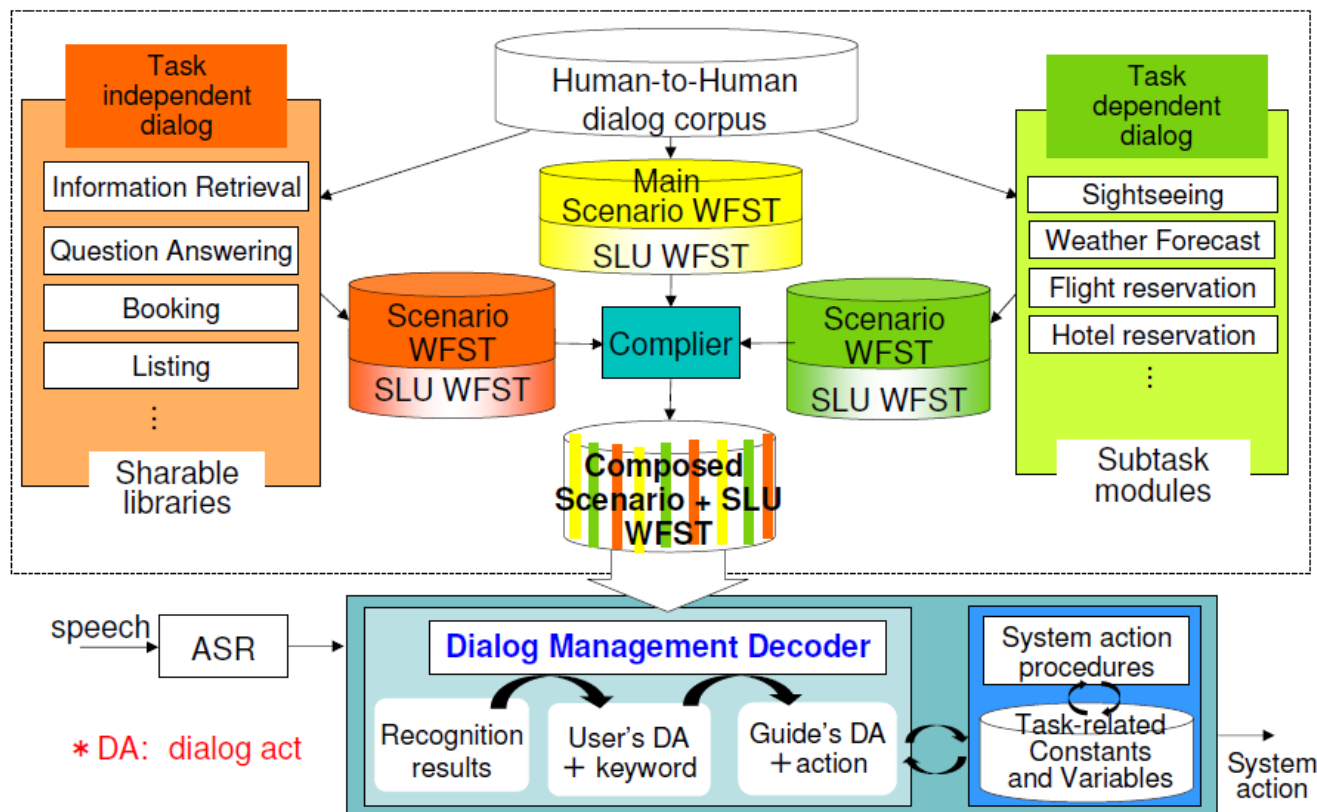
音声合成: HMM音声合成

- 多言語対応
 - 日本語、英語、中国語、韓国語、インドネシア語、マレー語、ベトナム語、ブラジルポルトガル語
- SSML入力
 - Speech Synthesis Markup Language (W3C) に準拠
 - 言語・声質・発話スタイルを文中で切り替えることができる
 - 間の長さ・挿入タイミングなどを制御可能(未対応の言語あり)
 - 対話用タグなど、NICT独自の拡張タグセットを追加(対話用タグは日本語のみ対応)
- Rubyベースのシステム
 - 生産性・保守性に優れる



統計制御モデルによる対話制御: WFSTDM

- 人の対話から学習し、対話制御モデルを構築
 - 統計的対話制御モデル WDSTを利用



インターネット上の情報などを活用し、文化的な背景や文脈を考慮した コミュニケーションを実現するナチュラル言語コミュニケーションの研究開発を進める

多言語で、多様な分野の情報を、多様な対象から、自然に取得するための基本技術開発

多様な分野で利用できれば

対象分野の拡張:

観光、医療、災害復興...

長い発話でも大丈夫かな？

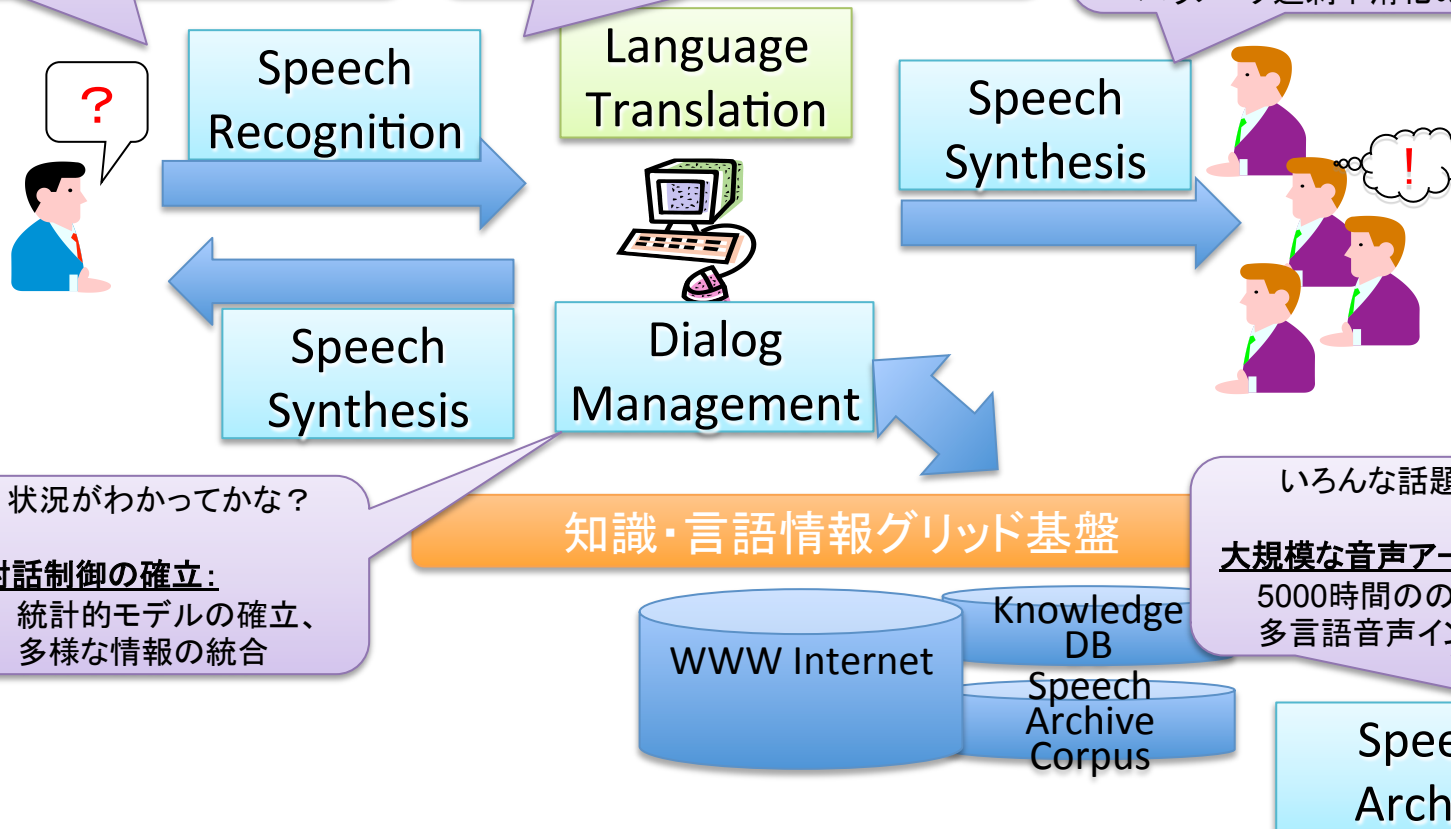
長文・高速音声認識:

言語モデル・認識アルゴリズムの改良

自然な発話をしてくれると！

高精度音声合成モデルの確立:

位相を考慮したモデル
パラメータ過剰平滑化の回避



状況がわかってかな？

対話制御の確立:

統計的モデルの確立、
多様な情報の統合

いろんな話題に対応してるかな？

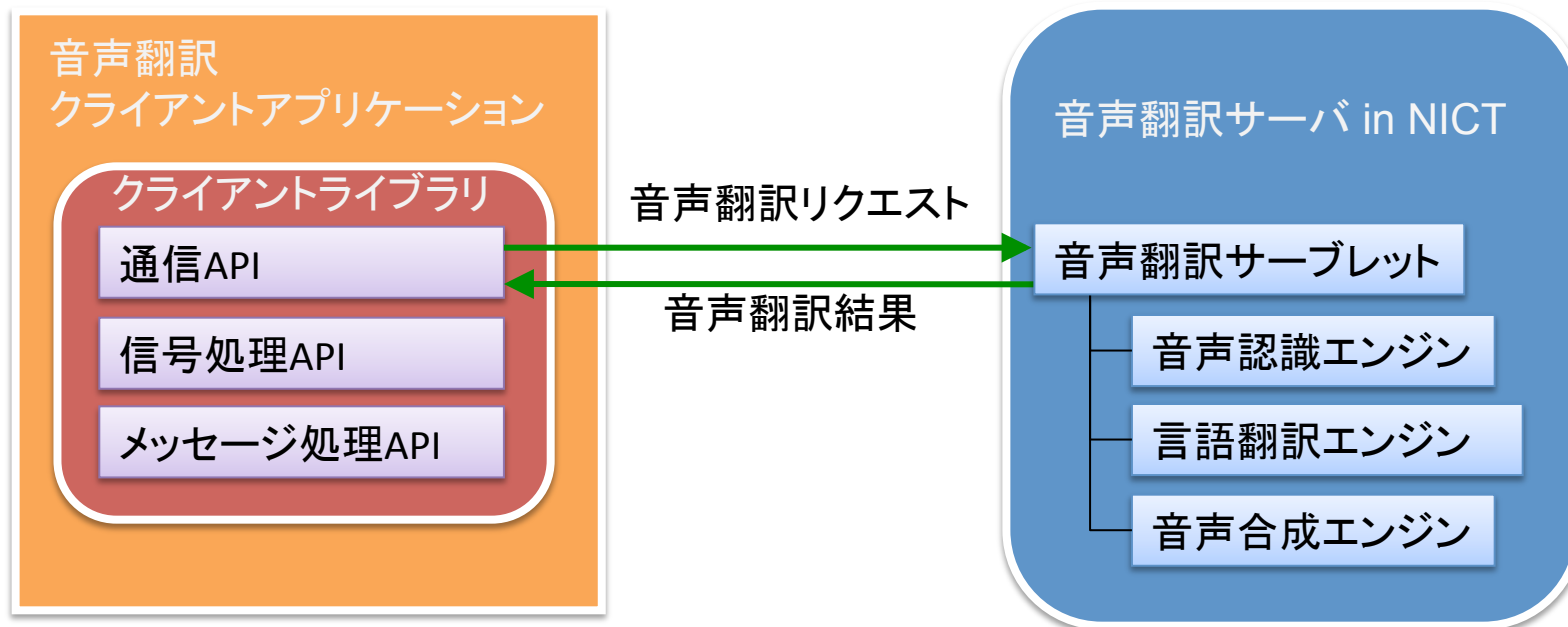
大規模な音声アーカイブ構築:

5000時間のコーパス(現状の5倍)
多言語音声インデキシング技術開発

音声翻訳ソフトウェア開発キット (SDK)

音声翻訳SDK

音声翻訳のクライアントアプリケーションを簡単に開発するためのソフトウェア開発キット。



特長

NICTがクライアントライブラリおよび試用サーバーを無償で提供するため、サーバー設置のための初期投資が不要で、参入障壁が低い。
→ 事業者が独自アプリ・新サービスのテストマーケティングを実施できる。

事業化

NICTから事業者にはソフトウェア(特許含む)を有償で供与する。
→ 事業者は、自らの投資でサーバーを設置し、商用サービスを行う。

音声翻訳SDKリリース内容

- 入手方法

音声翻訳SDK

検索

- クライアントライブラリ

iOS用のライブラリ

Android用のライブラリ

※ライブラリは、通信、信号処理、メッセージ処理APIを含みます。

- ドキュメント

iOS用のAPI仕様書

Android用のAPI仕様書

- サンプル

音声翻訳サンプルアプリ(iOS用、Android用)

ソースコード

仕様書

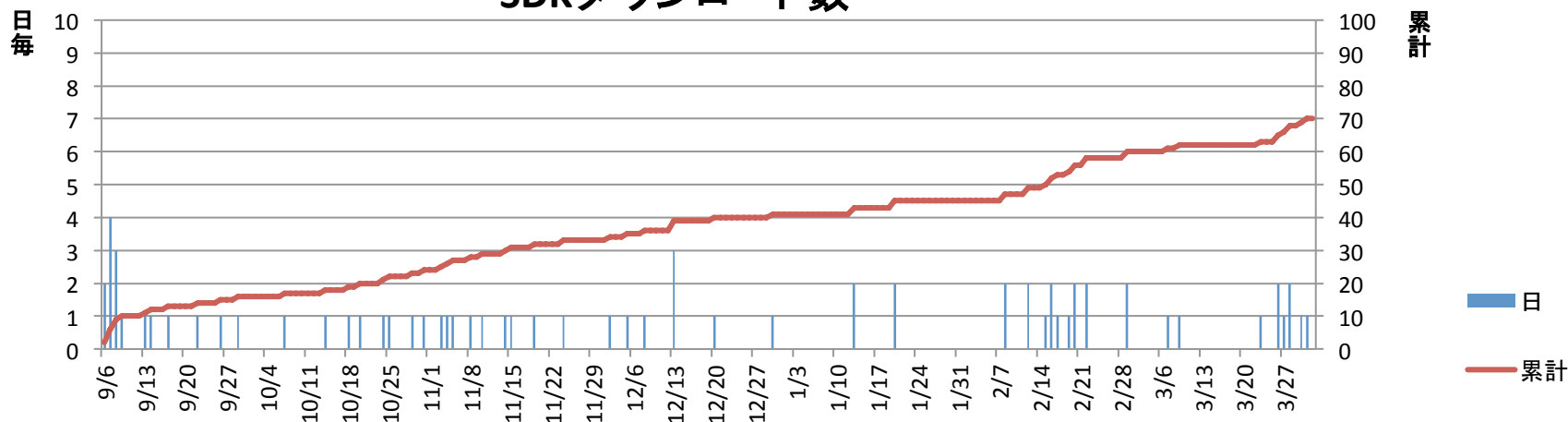
開発時の注意事項

SDK利用者向け制限事項

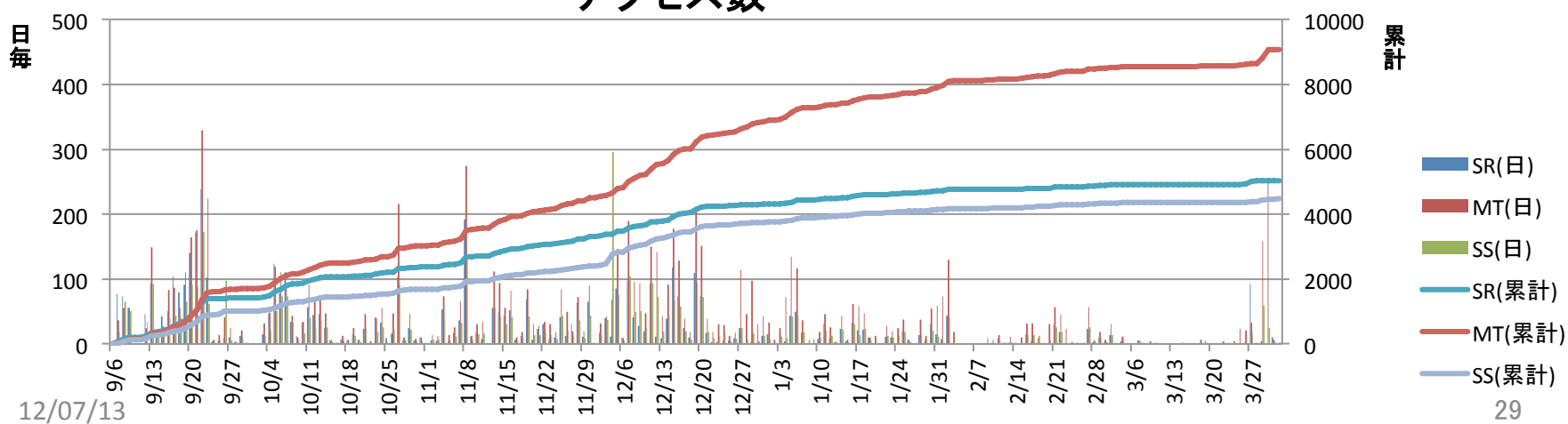
1. **商用利用**を目的とした又は**公序良俗に反する**アプリケーションで音声翻訳SDKを利用することはできません。商用利用の場合には別途有償契約が必要となります。
2. NICTは利用者による音声翻訳SDKの不適切な利用を発見した場合、予告無しに該当のアプリケーションからの音声翻訳サーバの利用を制限することがあります。
3. 利用者によるリクエスト頻度やサーバの繁忙によっては、NICTの任意の判断により、音声翻訳リクエスト回数に上限を設けることがあります。
4. NICTは予告無しに音声翻訳SDKの仕様を変更することがあります。NICTによる今後の音声翻訳SDKの提供・アップデート期間、**音声翻訳サーバの提供期間**は保証されません。
5. 音声翻訳のために入力された音声やその翻訳結果はNICTのサーバに記録されます。各記録は音声翻訳技術の改良のために利用される場合があります。ただし、入力音声や翻訳内容がそのまま公開されることはありません。

音声翻訳SDKの実績

SDKダウンロード数



アクセス数



京のおすすめ

- 気分と体験を組み合わせるなど、指定された複数の項目から、ユーザにぴったりの観光スポットを推薦します。



多言語音声コミュニケーション技術

頑健で分かりやすい音声処理

実環境下での高精度な音声認識 + コミュニカティブな音声合成 + 環境に応じた音響処理

◆ 多言語音声認識技術

実環境での認識性能の向上

多言語対応: 多言語対話音声の収録 / 非母語話者の音声収録

◆ 多言語音声合成技術

コミュニケーションな音声: 多様な対話様式への対応

多言語対応: 韻律情報の制御機構

人に優しい対話システム

的確に発話を理解した応答 + 非言語情報を使った息の合った応答 + 意図を酌んだ対応

◆ 対話制御技術の開発

対話モデルの構築

◆ コミュニケーションの評価

システムを介した実環境対話における情報伝達量を定量化

