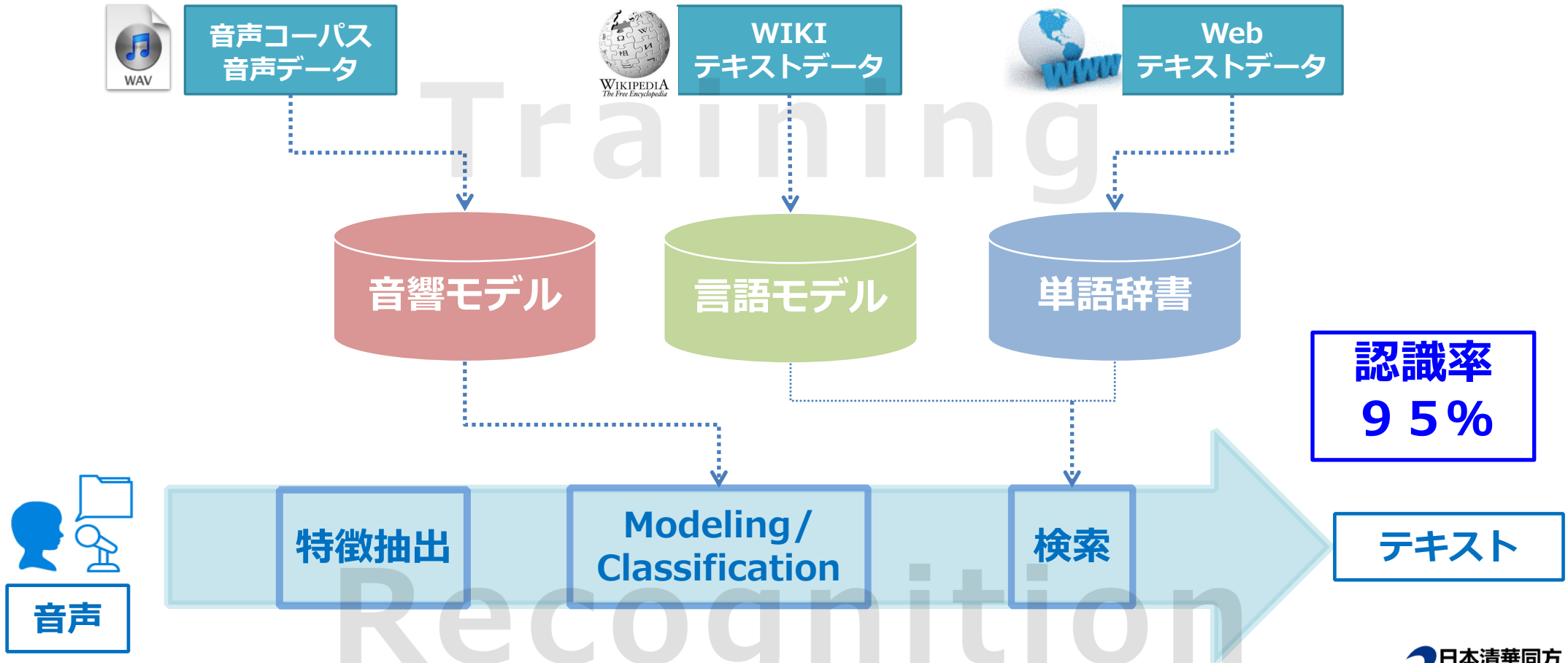
音声認識研究内容紹介

目次

- 音声認識の全体構成図
- 属性説明
- 音声認識エンジンの比較
- 知識のトランスファー（KT）技術説明
- 訓練用音声データの採取
- 運用イメージ

音声認識の全体構成図


音声コーパス
音声データ
WIKI
テキストデータ
Web
テキストデータ
Training
音響モデル
言語モデル
単語辞書
認識率
９５％
特徴抽出
Modeling/
Classification
検索
テキスト
音声
Recognition

属性説明

音響モデル	音声ドメイン	Deep Learning技術	モデルサイズ
AM	講演	○	76 MB

言語モデル	学習用データ		辞書	LMサイズ（MB）
	データ名	ドメイン		
LM	WEB 07	–	vocab-20170217	457
	WIKI	百科事典		

辞書	単語の数	選定基準
vocab	89556	WIKIに出現頻度が高い順（総数約9万まで）

デコーディング	サイズ（GB）	認識処理速度（1秒音声データに対する）
decode	6～7	0.45秒

音声認識エンジンの比較

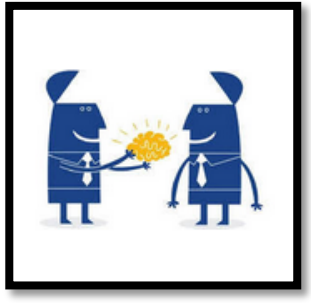
	IBM	Google	Microsoft	日本清華同方
開発元	IBM社	Google社	Microsoft社	清華同方研究新規開発
音声認識サービス運用	IBMクラウドにアップロード	Googleクラウドにアップロード	MSクラウドにアップロード	**お客様の自由 クラウド運用ができる オンプレもできる**
日本語音声認識	97% 個別再訓練対応しない	99% 個別再訓練対応しない	97% 個別再訓練対応しない	95.8% お客様要望より、業務向け強化再訓練が可能
ＡＭモデル訓練用データ資源量	数万時間	数万時間	数万時間	数百時間(＊KT技法)
サービス(I/F)	サービス公開（SDK）	サービス公開（SDK）	サービス公開（SDK）	SDK公開 WebService (Socket/NodeJS) お客様要望より、カスタマイズし、第三者に非公開も可能
Knowledge Transfer 技法	なし	なし	なし	中国語の音声認識ニューラルネットよりKT ※

※KT(Knowledge Transfer) は中国で特許提出済、現在は中国特許庁で審査中

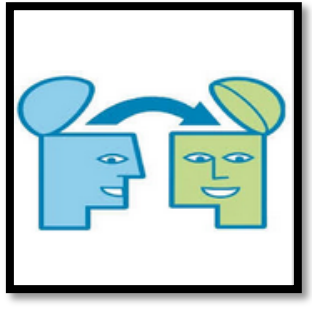
日本清華同方 JAPAN TSINGHUA TONGHUA

知識のトランスファー（KT）技術説明

- AIの転移学習技術です

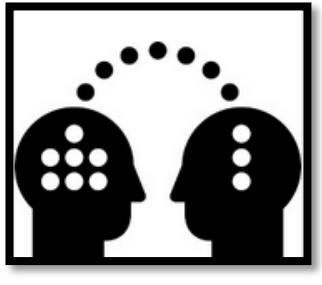


- AIはAIをトレーニングする

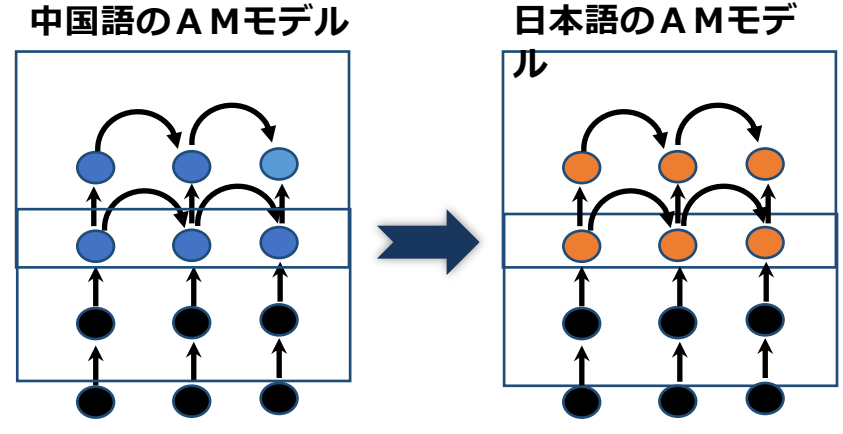


- 目的は学習データ準備の大変さを緩和

注：中国語の音声資源は5000H超、それらを強い計算環境でAM生成済み、ちなみに日本語の資源は660H

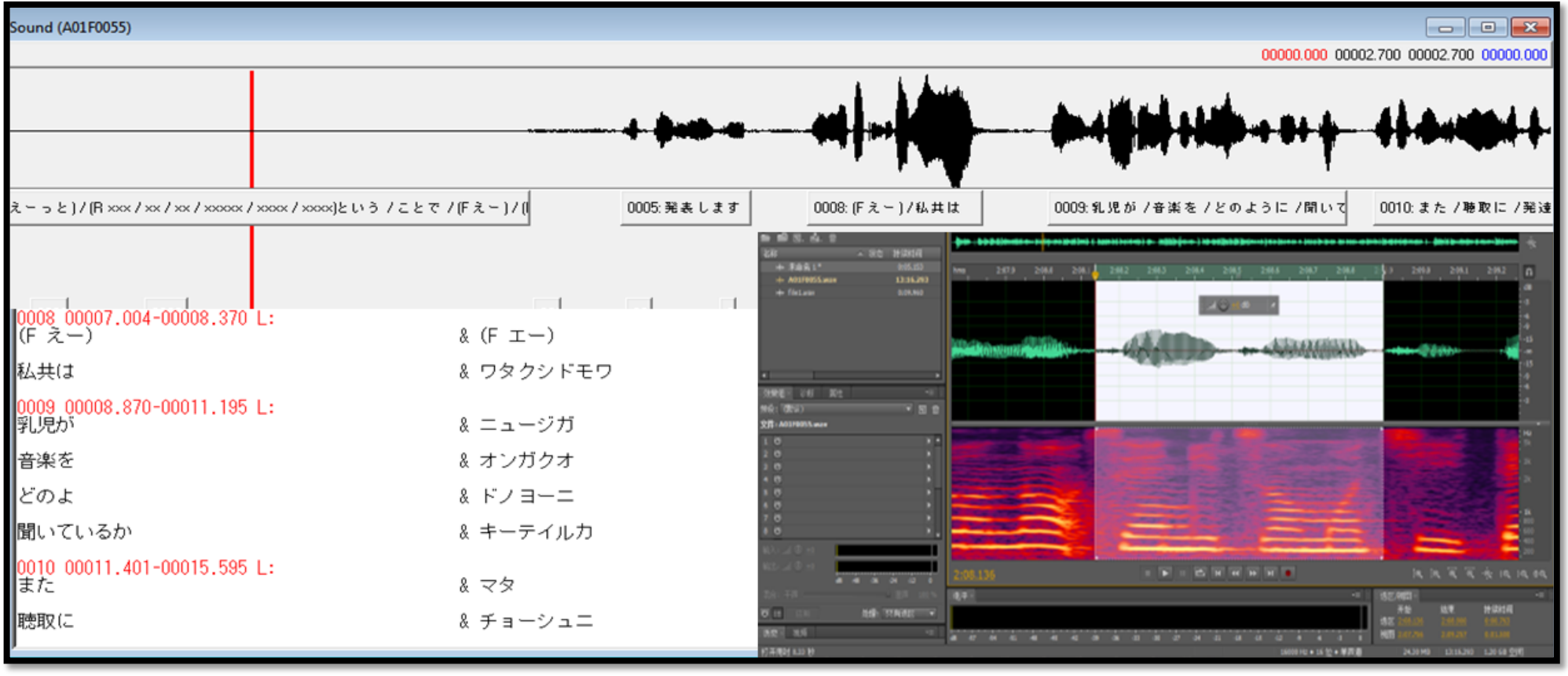


- 方言の音声認識時にモデルの作成に活用できる



講師ありのDNN（深層ニューラルネット）学習手法の下で神経細胞間と階層間の情報伝授にKTの介入で、パラメータの修正を加えもっとも有効かつ正確な調整は技術で資源データの不足と計算能力不足の不足を大きな意味で補えた。

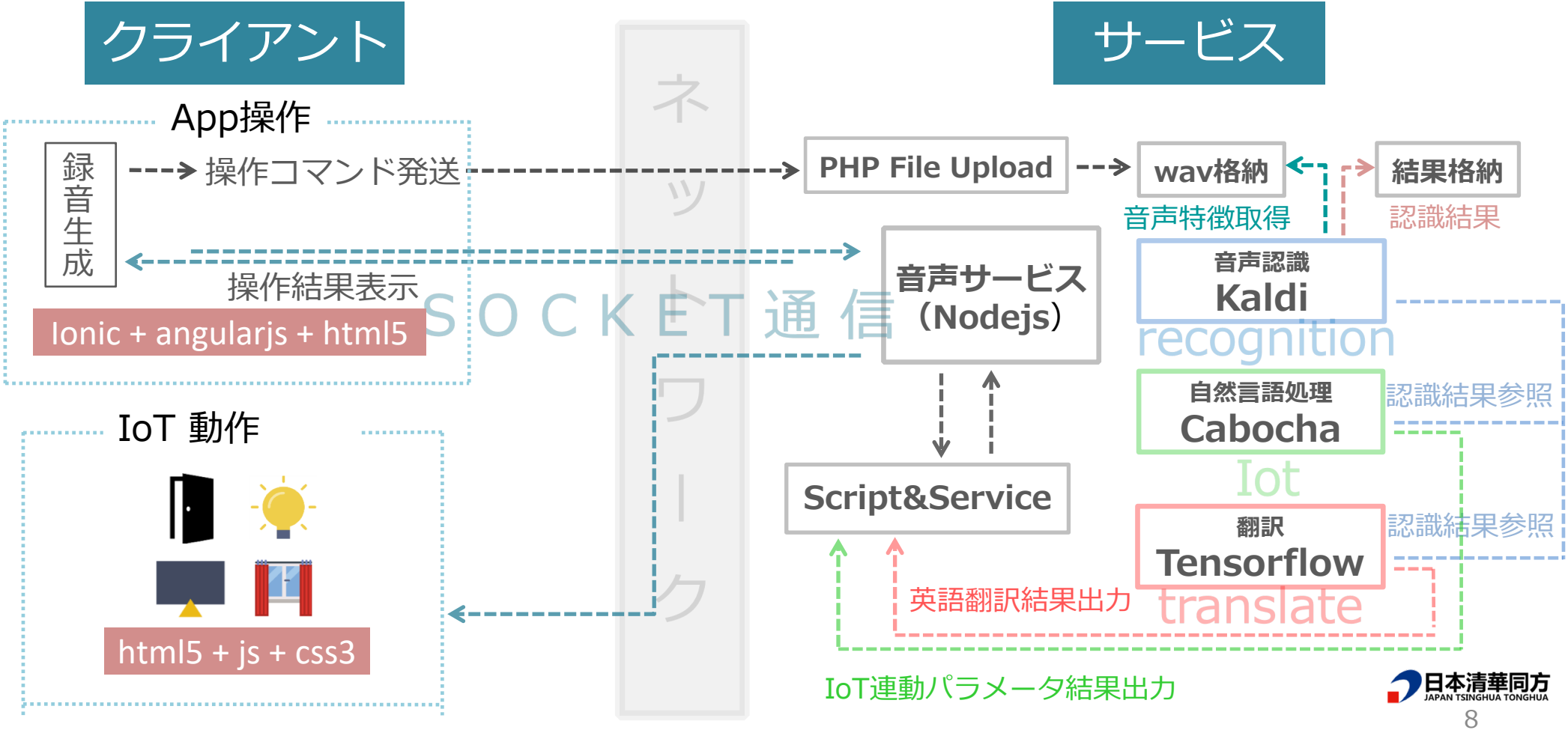
訓練用音声データの採取：区切りとラベリング※



※NN深層学習の性格上、大量に採取できた音声は正しくマックつけ、講師ありの訓練材料となる

日本清華同方 JAPAN TSINGHUA TONGHUA

運用イメージ


クライアント
サービス
App操作
録音生成
操作コマンド送信
操作結果表示
Ionic + angularjs + html5
ネットワーク
SOCKET通信
PHP File Upload
wav格納
結果格納
音声特徴取得
認識結果
音声サービス
(Nodejs)
音声認識
Kaldi
recognition
自然言語処理
Cabocha
認識結果参照
Iot
Script&Service
翻訳
Tensorflow
認識結果参照
英語翻訳結果出力
translate
IoT 動作
html5 + js + css3
IoT連動パラメータ結果出力

ご清聴ありがとうございました

日本清華同方
JAPAN TSINGHUA TONGHUA