



Speech-to-Speech Translation Services for the Olympic Games 2008

Sebastian Stüker, Chengqing Zong, Jürgen Reichert,
Wenjie Cao, Muntsin Kolss, Guodong Xie, Kay
Peterson, Peng Ding, Victoria Arranz, Jian Yu, Alex
Waibel

Introduction

- Special Programme for Construction of Digital Olympics
- Overview over the STT System
- Domains Adressed
- System Integration
- Single Components
- Hardwarer Setup
- User Interface



Digital Olympics 2008

- In 2008 the Olympic Summer Games will be held in Beijing, China
- Beijing city government launched the **Special Programme for Construction of Digital Olympics**
- One of the objectives: „... to remove language barriers with the aid of artificial intelligence technology in order to promote friendship and mutual understanding“
- Our contribution: Rapidly develop a speech-to-speech translation prototype to demonstrate the feasibility of deploying translation technology in 2008 in Beijing



DO Partners



**Universität
Karlsruhe (TH)**



Carnegie Mellon

MOBILE TECHNOLOGIES L.L.C.

Speech-to-Speech Translation Services for the Olympic Games 2008
Sebastian Stüker, Universität Karlsruhe (TH)



System Overview

- For STT three technologies involved:
 - Automatic Speech Recognition (EN, CH, SP)
 - Machine Translation (Interlingua, 2x SMT)
 - Speech Synthesis
- Component Integration
 - Client Server Setup
 - Active Speech Frame Work



Domains Adressed

- Domain Limited System
- Primary Domain: Pre-Arrival Hotel Reservation:
 - **Not** a face-to-face scenario
 - **But** prior project experience and available resources
- Secondary Domains:
 - Basic tourist needs (BTEC)
 - Basic medical needs (provides leverage)



Active Speech Framework

- Environment for building and testing multi-modal interfaces
- Allows for easy creation of demos and prototypes, and analysis of interface issues
- Based on the dataflow paradigm:
 - Each component receives and transforms data
 - Sends the result to the next component that transforms and sends again etc.



ASR Components

ASR Engines for three languages:

- **English (UKA):**
 - Janus Recognition Toolkit feat. IBIS decoder
 - Derived from the ISL 2004 Meeting Recognition System
 - 2.5K vocabulary, 3-Gram LM with semantically defined classes (names, points of interest, etc.)
 - Incremental speaker adaptation (VTLN, MLLR, fMLLR)
 - 18.1% WER on hotel reservation dialogs
- **Spanish (UKA):**
 - Janus Recognition Toolkit feat. IBIS decoder
 - 24k vocabulary, 3-Gram LM
 - Incremental speaker adaptation (VTLN, MLLR, fMLLR)
 - 16.3% WER on hotel reservation dialogs
- **Chinese (NLPR):**
 - 50k vocabulary, interpolated 4-Gram and class based 3-Gram model
 - In-house, time-synchronous triphone decoder, special consideration of tonal information of context-phones



MT Components

Interlingua Based (CMU, NLPR)

- Between English, Spanish, Chinese
- Analyse into a high level Interlingua (represents semantics), then generate from Interlingua into target language
- Covers the hotel reservation domain

Statistical Machine Translation (CMU, UKA)

- Between English, Spanish, Chinese
- Based on the noisys channel approach
- Trained on BTEC and in-house collection of hotel reservation dialogs
- ISL SMT system (direkt, bilingual transfer) and EDTRL (enriched English as Interlingua)



Phrasebook

- Provides a fast repair mechanism in case of ASR errors
- Selects similar phrases by means of editing distance
- Phrasebook contains 160k phrases from BTEC and 100k translation dictionary
- BTEC Phrases were classified into 13 classes by NLPR, so repair phrases can be constraint manually



Synthesis Component

English and Spanish:

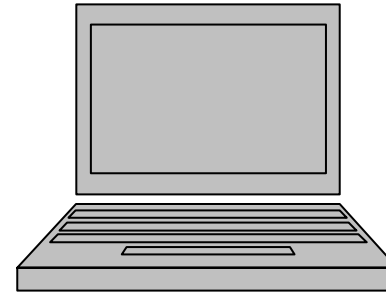
- provided by Cepstral LLC, Pittsburgh, PA
- Unit selection based
- Small enough to run on small portable devices (e.g. PDAs)

Chinese:

- Provided by NLPR
- Concatenative
- Small enough to run on small portable devices (e.g. PDAs)



Hardware Setup



Server-Laptop:
ASR, MT, Synthesis

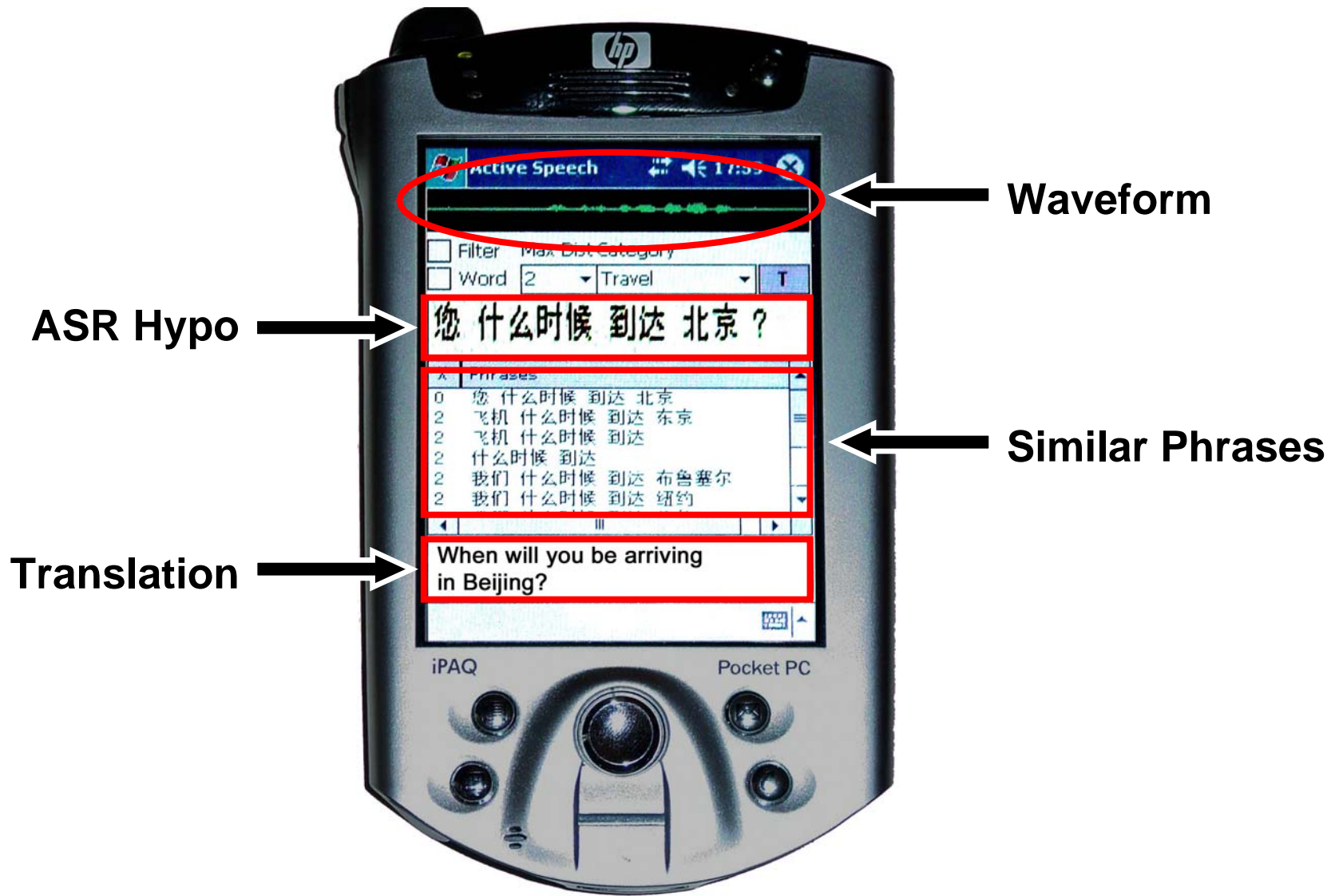


PDA-Clients:

- Push-To-Talk
- present results



User Interface



User Interface - Laptop

Component Selection

ASR Hypo

Similar Phrases

Translation

Component Selection

Component Control

Waveform

Component	Address	Started
AS_Map		
AudioIn		
AudioOut	141.3.25.136:4791	started
CMU_GenKitMapp...	141.3.25.136:4768	started
CMU_ParseMapp...	141.3.25.136:4792	started
CMU_ParseMapp...		
CMU_SoupE	141.3.25.136:4766	started
CMU_SoupS		
DllExample		
Dotnet Example		
EDTRL_CE	141.3.25.136:4771	started
EDTRL_EC	141.3.25.136:4770	started
FileExample		
FileIn		
FileOut		
GPS		
JanusC		
JanusCL		
JanusCWilson	141.3.25.136:4773	started
JanusE_8KHz_MT		
JanusE_MT	141.3.25.136:4781	started
JanusE_small		
JanusS		
Logger	141.3.25.136:4761	started
NLPR_ASR	141.3.25.136:4767	started
NLPR_NLG	141.3.25.136:4782	started
NLPR_PARSE	141.3.25.136:4803	started
NLPR_TTS	141.3.25.136:4795	started
PY2CW		
SAPL_TTS_C_MS		
SAPL_TTS_E	141.3.25.136:4776	started
SAPL_TTS_E_SAM		
SAPL_TTS_S_Rocio		
SAPL_TTS_Thai		
SMT_CE	141.3.25.136:4778	started
SMT_CS		
SMT_EC	141.3.25.136:4787	started
SMT_ES		
SMT_SC		
SMT_SE		
Socket Example		
Test_IF		
Test_Translator		

X	Phrase	Translation
0	北京旅行社您需要什么吗	beijing travel agency how can i help yo
1	北京旅行社您需要什么	beijing travel agency how can i help yo

X	Phrase	Translation
0	hi i need a hotel room in beijing	您好我需要一个在北京的宾馆

Beijing Travel Agency .
May I help you ?

你好。



Conclusion



- Addressed a way of bridging the language barrier in human communication
- Developed Demonstration / Prototype Application
 - Integrates ASR, MT, TTS in a server-client setup
 - Allows for fast tailoring of demo systems
 - Supports direct comparison of different implementations of one component
- Showed the feasibility of Deploying SST technology in the environment of the Olympic Games 2008

