



TRANSONIC

ترجمه بلادرنگ گفتار به گفتار
فارسی - انگلیسی

زهرا منصوری

z_mansoori@ce.sharif.edu

فهرست

- تاریخچه ..
- معرفی ..
- معماری سیستم ..
- نتایج ..
- مشاهده نحوه عمل سیستم ..
- منابع ..

.. به سفارش آژانس فدرال ایالات متحده

☒ هدف: دست یافتن به فن آوری هایی با قابلیت استفاده چند ملیتی و چند زبانی

.. پیشنهاد پروژه جاری از سوی DARPA و سایر گروه های تحقیقاتی DoD

.. پیاده سازی توسط دانشگاه Southern California

معرفی Transonic

4

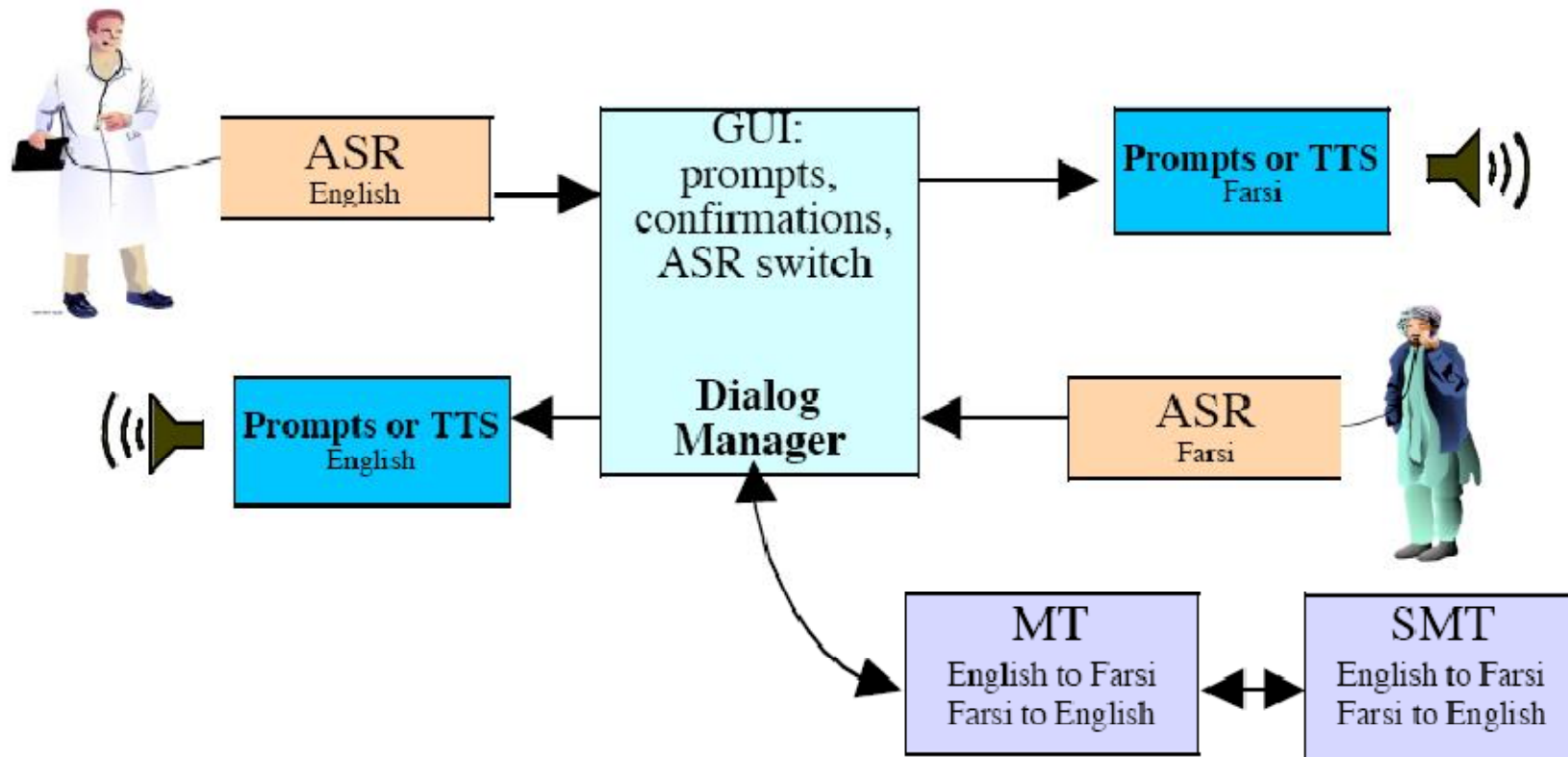
- سیستمی ارتباطی که میان کاربران تک زبانه که به یکی از دو زبان فارسی و انگلیسی تکلم می کنند، ارتباط کلامی برقرار می کند
- برای موارد خاص پزشکی-درمانی طراحی شده است
 - ✘ پزشک معالج به زبان انگلیسی
 - ✘ بیمار به زبان فارسی

اجزای تشکیل دهنده Transonic

- .. بخش پردازش خودکار گفتار (ASR)
- .. بخش مدیریت دیالوگ ها (DM)
- .. بخش ترجمه ماشینی (MT)
- .. بخش سنتز متن به گفتار (TTS)
- .. واسط گرافیکی کاربر (GUI)

اجزای تشکیل دهنده Transonic

6



عملکرد Transonic

۱. در ابتدا صوت ورودی توسط یک واسط دریافت صوتی (Audio Client) دریافت می شود
۲. گفتار دریافت شده با استفاده از بخش ASR به متن معادل تبدیل می شود
۳. متن خروجی با استفاده از بخش MT ترجمه می شود
۴. متن ترجمه شده در TTS دوباره به صوت تبدیل شده و توسط واسط پخش صوت (Play-out Client) پخش می شود.

بخش پردازش خودکار گفتار (ASR)

Automatic Speech Recognition

۱. صوت ورودی را با استفاده از **Audio Client** دریافت می کند
۲. لیستی از متن های معادل با صوت ورودی را به دست می دهد
این لیست مشتمل بر n متن است که در بین تمامی متون معادل با صوت ورودی، از درجه صحت بالاتری برخوردارند
- .. **ASR** انگلیسی با استفاده از ۲۲۰۰۰ لغت و **ASR** فارسی با استفاده از ۹۰۰۰ لغت کار می کند
- .. به دلیل کاربری خاص این سیستم، آموزش **ASR** مطلوب آسان تر از حالت کلی است

بخش مدیریت دیالوگ ها (DM)

Dialog Manager - به عنوان قلب سیستم

۱. همه پیام ها با برجسبی که شامل اطلاعات مربوط به میزان داده، مبدا و مقصد پیام است برجسب می خورند
۲. این پیامها توسط DM به تمامی کامپوننت ها Broadcast می شوند
۳. نتایج خروجی از ASR را بر اساس درجه صحت آن ها قبول یا رد می کند
۴. از طریق واسط گرافیکی از کاربر درباره صحت ترجمه پرس و جو می کند
۵. کنترل ارسال و مونیتورینگ پیام ها را برعهده دارد
۶. توانایی متوقف کردن کار کامپوننت ها را نیز داراست (برای پرس و جو از کاربر، نظیر تکرار، تایید عبارت و رفع ابهام آن از طریق واسط گرافیکی)

بخش ترجمه ماشینی (MT)

Machine Translation ..

ترجمه مکالمات بین دو شخص را برعهده دارد ..

ترجمه به دو صورت انجام می گیرد ..

۱. ترجمه آماده متون دخیره شده

۲. ترجمه آماری با استفاده از SMT (Statistic Machine Translation)

انتخاب روش ترجمه با استفاده از یک کلاسبندی کننده انجام می گیرد ..

به دلیل کاربری محدود این سیستم، استفاده از کلاسبندی متون ورودی میسر است ..

نحوه کار کلاسبندی کننده

در مرحله آماده سازی سیستم

- تعدادی از مفاهیم مرتبط با مکالمات پزشکی و ترجمه معادل آن ها در سیستم ذخیره می شود
- این مفاهیم به کلاس هایی تقسیم بندی می شود
- در هر کلاس تعدادی متن معادل با مفهوم کلاس ذخیره می شود

نحوه کار کلاسبندی کننده - ادامه

در مرحله اجرا

•• کلاسبندی کننده متن ورودی را دریافت می کند

•• اگر میزان شباهت مفهوم از حد آستانه بیشتر بود

✘ متن ورودی را به یکی از کلاس های موجود در سیستم نسبت می دهد

✘ از ترجمه ذخیره شده آن مفهوم (کلاس) به عنوان ترجمه عبارت ورودی استفاده می شود

•• در غیر این صورت

✘ از ترجمه آماری (لغت به لغت و با استفاده از مدل زبانی پزشک یا بیمار) برای

ترجمه متن ورودی استفاده می شود

واسط گرافیکی کاربر

•• برای افزایش دقت ترجمه ماشینی مکالمه پزشک-بیمار در دو حالت استفاده می شود

•• سه نوع کاربری

۱. تایید صحت ترجمه گفتار پزشک برای اعلام به بیمار
۲. تایید صحت ترجمه گفتار بیمار
۳. مشاهده تاریخچه مکالمه

واسط گرافیکی کاربر - ادامه

۱. تایید صحت ترجمه گفتار پزشک

- ✘ برای مثال، ۵ گزینه ترجمه گفتار بیمار برای پزشک به نمایش در می آید
- ✘ یکی با استفاده از روش آماری و ۴ ترجمه دیگر با استفاده از کلاسبندی به دست می آید

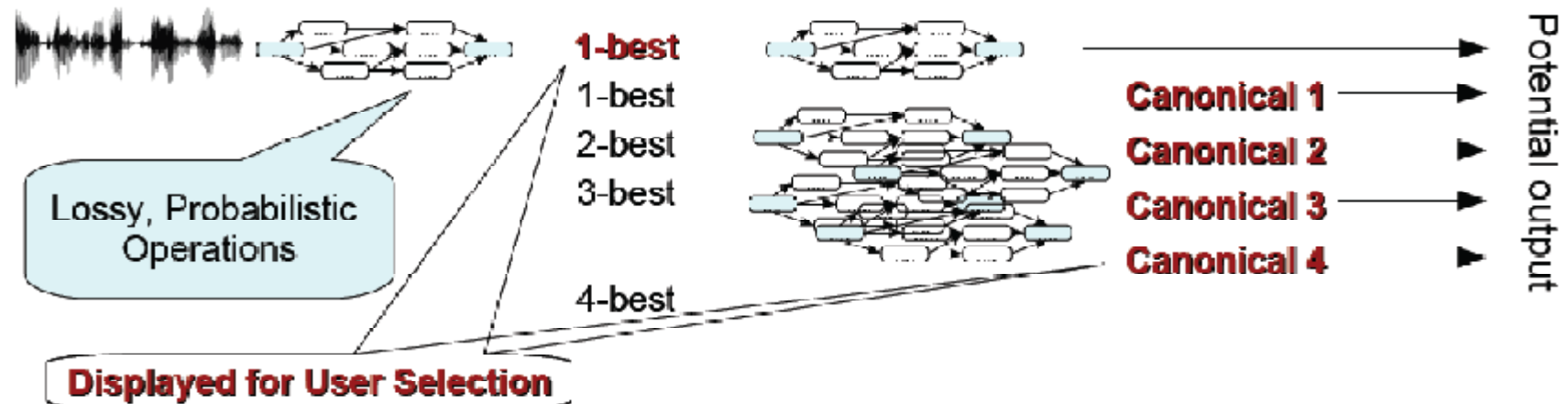
- ✘ این ترجمه ها به ترتیب میزان صحت آن ها به پزشک نمایش داده می شوند
- ✘ پزشک بهترین ترجمه گفتار بیمار را انتخاب می کند

۲. تایید صحت ترجمه گفتار بیمار

- ۱. به شیوه قبل، گفتار بیمار به ۵ صورت ترجمه می شود
- ۲. پزشک بر اساس تشخیص خود و تاریخچه مکالمه، بهترین و نزدیکترین گزینه را انتخاب می کند

واسط گرافیکی کاربر

15



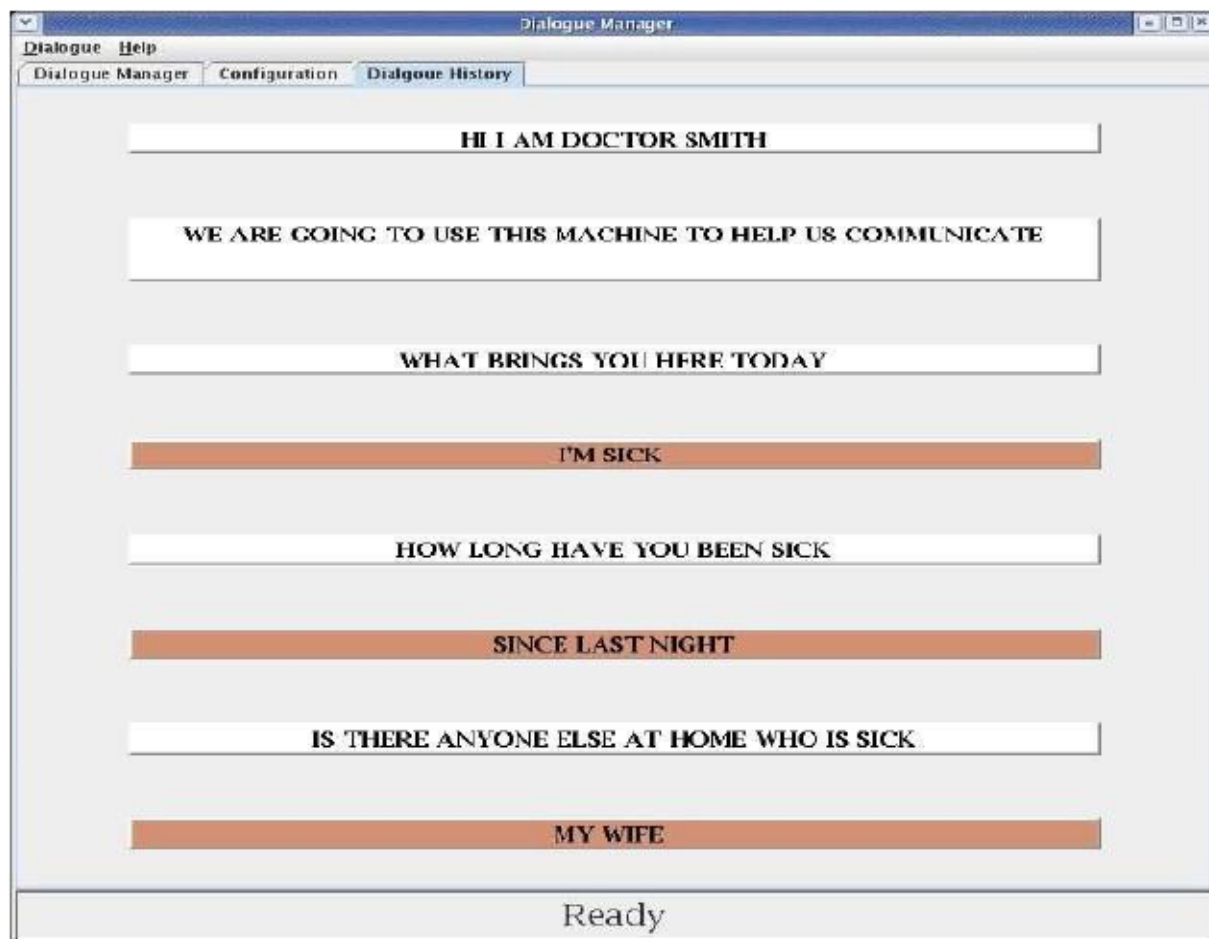
نحوه انتخاب بهترین ترجمه گفتار پزشک

16



نمایش تاریخچه مکالمه

17



سنتز متن به گفتار (TTS)

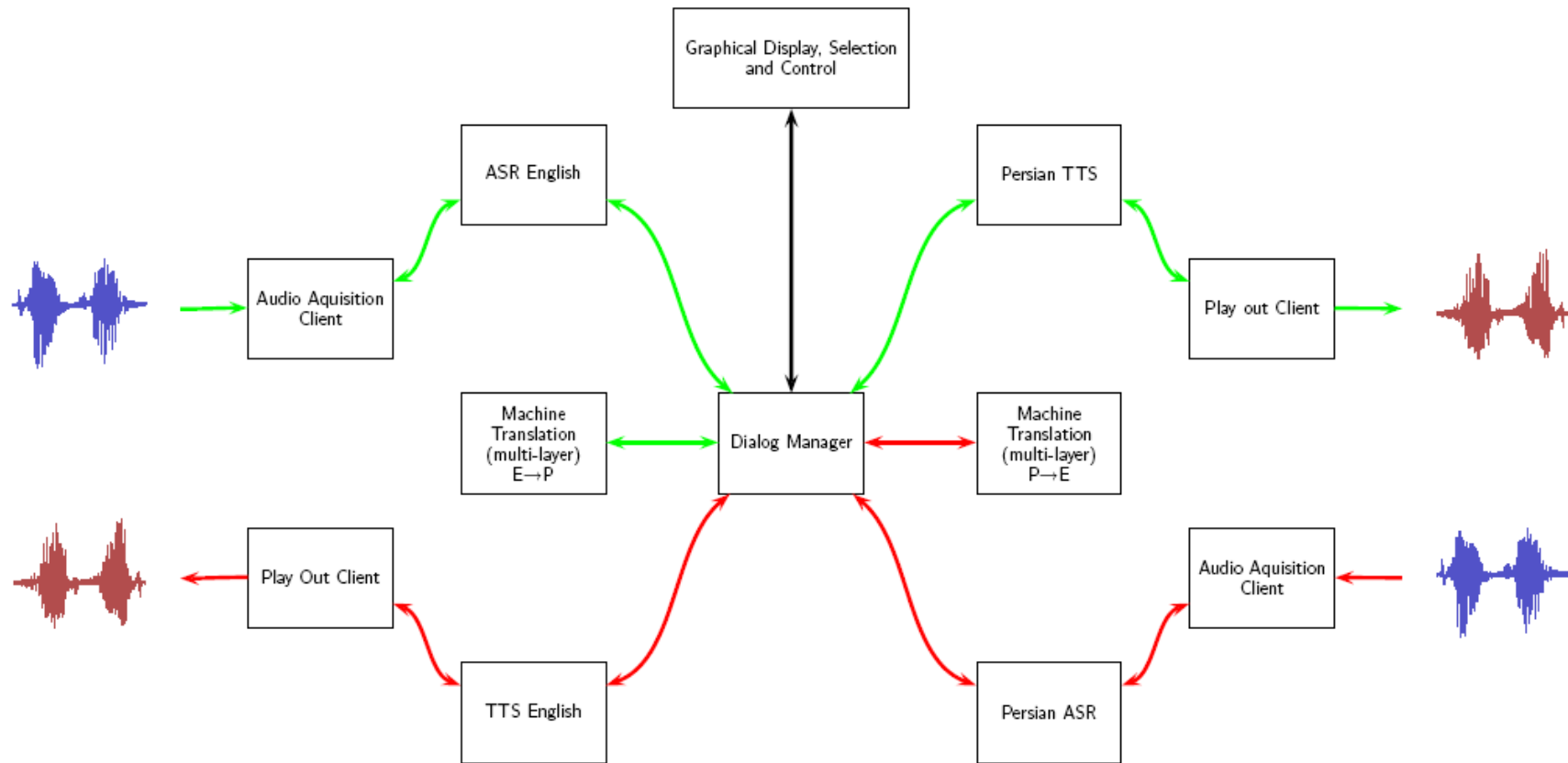
Text To Speech ..

.. متن ترجمه شده را به صوت تبدیل می کند

.. چندین حالت کاری دارد:

۱. برای عباراتی که در حوزه دانش کلاسبندی کننده قرار می گیرند، صدای طبیعی یک فرد که از پیش ضبط شده است پخش می شود.
۲. اگر عبارت به صورت آماری ترجمه شده باشد، واحد سنتز از سطح عبارات سطح کلمات نزول می کند و سلسله ای از گفتار های از پیش ضبط شده برای آن کلمات پخش می شوند.
۳. برای کلماتی که صورت از پیش ذخیره شده ای برای آن ها وجود ندارد، یعنی لغات غیر معمول، سنتز در سطح **diphone** انجام می شود.

ارتباطات پیامی بین کامپوننت ها



معماری سیستم

- .. در معماری جاری، تمام سیستم ها به صورت شبکه قابل اتصال هستند
- .. بنابراین نیاز به جمع آوری آن ها در یک سیستم واحد نیست
- .. همگی این اجزاء قابل اجرا بر روی سیستم عامل **Windows** و **Linux** هستند.

پیاده سازی ASR

برای ASR انگلیسی

- از پردازنده Sonic دانشگاه Colorado استفاده شده است
- از ابتدا توسط دیالوگ های پزشک-بیمار جمع آوری شده از دانشجویان پزشکی-آموزش دیده است

برای آموزش ASR فارسی

- به دلیل کمبود داده های صوتی از مدل صوتی انگلیسی با استفاده از یک نگاشت Sub-phonetic بین آواهای انگلیسی-فارسی ساخته شده است
- استفاده از مجموعه اصوات فارسی به نام FARSDAT
- استفاده از اصوات گفتاری افراد فارسی زبان

پیاده سازی MT

22

بخش MT انگلیسی

•• از ۱۴۰۰ کلاس مختلف برای زبان انگلیسی

•• هر کلاس

✘ حاوی عبارات متداول و استانداردی است که در محاوره های پزشک-بیمار استفاده می شود

✘ دارای مجموعه بسیار بزرگی از تفاسیر معادل با عبارات داخل کلاس می باشد

✘ اگر پزشک یکی از آن تفاسیر را به کار ببرد، سیستم عبارت معادل آن را تشخیص دهد و آن را با استفاده از **Lookup Table** به فارسی ترجمه کند

سیستم Transonic با دارا بودن پایگاه داده شامل ۳۰۰ هزار کلمه، به شیوه های زیر آزمایش شده است

•• ۱۵۰ میلیون کلمه مجزا که از Web به دست آمده است و برای حوزه کاری سیستم فیلتر نشده است

•• ۱۵۰ میلیون کلمه مجزا از Web که برای حوزه کاری سیستم فیلتر شده است

•• پایگاه داده PPL

•• پایگاه داده BLEU

•• پایگاه داده LPU

خطای تشخیص در سیستم Transonic

	10K	20K	30K	40K
No Web	19.8	18.9	18.3	17.9
All Web	19.5	19.1	18.7	17.9
PPL	19.2	18.8	18.5	17.9
BLEU	19.3	18.8	18.5	17.9
LPU	19.2	18.8	18.5	17.8
Proposed	18.3	18.2	18.2	17.3

- .. سیستم برای داده های حاصل از وب که برای این حوزه کاری فیلتر نشده اند، بازده کمتری را نشان می دهد
- .. این به معنای وابستگی تشخیص سیستم به حوزه کاری محدود آن شده است
- .. بر اساس مشاهده بازده کاری این سیستم قابل قبول است
- .. به گفته توسعه دهندگان این سیستم، نتیجه فاجعه بار (تشخیص نادرست و زیانبار بیماری) توسط این سیستم گزارش نشده است

- 1) Emil Ettelaie, *"Cross-lingual Dialog Model for Speech to Speech Translation"*, ICSLP, INTERSPEECH 2006
- 2) Shrikanth Narayanan, *"Speech Recognition Engineering Issue In Speech To Speech Translation System Design For Low Resource Languages And Domains"*, IEEE, 2006
- 3) JongHo Shin, *"User Modeling in a Speech Translation Driven Mediated Interaction Setting"*, Viterbi School of Engineering, 2006
- 4) Emil Ettelaie, *"Transonics: A Practical Speech-to-Speech Translator for English-Farsi Medical Dialogues"*, Proceedings of the ACL Interactive Poster and Demonstration Sessions, June 2005
- 5) S. Narayanan, *"The Transonic Spoken Dialogue Translator: An Aid For English-Persian Doctor-Patient Overviews"*, University of Southern California, 2004
- 6) Panayiotis G. Georgiou, *"An English-Persian Automatic Speech Translator: Recent Developments In Domain Portability And User Modeling"*, University of Southern California
- 7) <http://sail.usc.edu/transonics/srs.php>