به نام خدا

Transonic:

ترجمه‌ی بلادرنگ گفتار به گفتار

انگلیسی- فارسی

گردآوری: زهرا منصوری

استاد: دکتر ثامتی

دانشکده مهندسی کامپیوتر

تابستان ۸۶

فهرست

۱- چکیده

هدف از این گزارش، ارائه تعریفی از ساختار و نحوه عملکرد سیستم‌های ترجمه گفتار به گفتار از زبان انگلیسی به فارسی و بالاعکس به منظور ارتباط پزشک و بیمار می‌باشد، به صورتی که پزشک و بیمار به یک زبان تکلم نمی‌کنند. با توجه به این موضوع میزان دقت عملکرد سیستم بسیار با اهمیت خواهد بود. برای بالا بردن این دقت تنها به بهبود بازده

اجزای استاندارد تشکیل دهنده به صورت منفرد اکتفا نمی شود، بلکه سعی بر آن است تا بازده سیستم به طور کلی افزایش یابد. با توجه به محدود بودن حوزه کاربری این سیستم، این امر به نجو موثرتری امکان‌پذیر خواهد بود.

۲- تاریخچه

آژانس فدرال ایالات متحده، در جستجوی دست یافتن به فن آوری‌هایی است که قابلیت استفاده چند ملیتی و چند زبانی را داشته باشند. یکی از چالش‌های مهم در این زمینه، توسعه‌ی سیستم‌هایی است که ارتباط بین کاربران با ملیت-های مختلف و زبان های متفاوت را برقرار می کنند. با توجه به این موضوع، DARPA و سایر گروه های تحقیقاتی DoD پروژه‌ای پیشنهاد کردند که ارتباط زبانی را با استفاده از ترجمه ی کامپیوتری گفتگو بین کاربران با زبان‌های مادری مختلف را برقرار کند. دانشگاه South California نمونه اولیه‌ای تحت نام راه‌کار Transonic را توسعه داده است. این راه‌کار ارتباط بین دو شخص با زبان‌های متفاوت را با استفاده از ترجمه دو طرفه گفتگو بین آن‌ها در دامنه زبان های انگلیسی – فارسی و انگلیسی- دری را برقرار می سازد.

۳- معرفی

Transonic سیستمی ارتباطی است که میان کاربران تک زبانه که به یکی از دو زبان فارسی و انگلیسی تکلم می-کنند، ارتباط کلامی برقرار می‌کند. این سیستم برای موارد پزشکی-درمانی طراحی شده است، به گونه ای که پزشک معالج به زبان انگلیسی و بیمار به زبان فارسی تکلم می کند.

این سیستم از هفت بخش پردازش زبان یا پردازش گفتار تشکیل شده است (شکل ۱). تمامی بخش ها توسط یک سیستم انتقال پیام به یکدیگر مرتبط‌اند.

اجزای تشکیل دهنده سیستم عبارتند از:

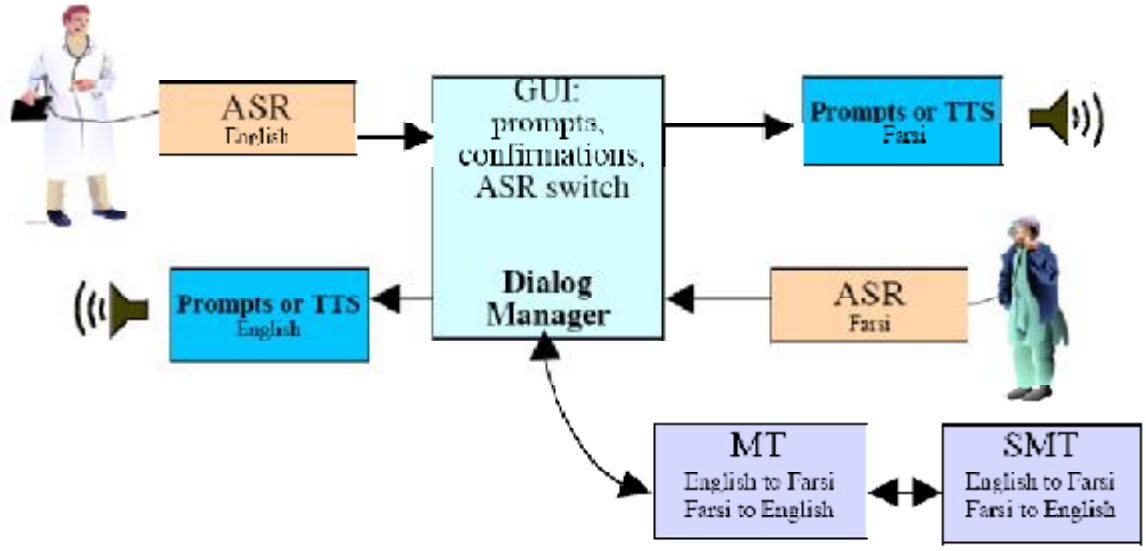
۱. بخش پردازش خودکار گفتار[۱] (ASR)
۲. بخش مدیریت دیالوگ ها (DM)
۳. بخش ترجمه ماشینی (MT)
۴. بخش سنتز متن به گفتار (TTS)
۵. واسط گرافیکی کاربر (GUI)

Transonic به این صورت عمل می‌کند که گفتار ورودی کاربر را تشخیص[۲] و آن را به متن تبدیل می‌کند. سپس متن ترجمه شده دوباره سنتز می شود تا به زبان مطلوب ثانویه ترجمه شود.

به طور مختصر، در ابتدا صوت ورودی توسط یک واسط دریافت صوتی[۳] (AC) دریافت می‌شود، سپس با استفاده از بخش ASR سیستم به متن معادل تبدیل شده و با استفاده از MT ترجمه می‌شود. این متن ترجمه شده در TTS دوباره به صوت تبدیل شده و توسط واسط پخش صوت[۴] پخش می‌شود.

در ادامه اجزای سیستم و وظایف هر یک از آن ها توضیح داده شده است.

[۱] Automatic Speech Recognition
[۲] Recognize
[۳] Audio Client
[۴] Play-out Client



شکل ۱ معماری سیستم Transonic

۳-۱- بخش پردازش خودکار گفتار (ASR)

زیر سیستم[۵] ASR گفتگوی ورودی را دریافت کرده و بر اساس آن لیستی از متون متناظر آن را ارائه می‌دهد که شامل صورت‌های مختلف حاصل از کدبرداری[۶] صوت ورودی است. در این لیست، n صورت از گفتگوی رونویسی شده[۷] موجود است که این در میان سایر صورت‌های رونویسی شده دارای درجه اطمینان اداراکی بالاتری می‌باشد.

این زیر سیستم به صورت بلادرنگ و در دو زبان انگلیسی و فارسی کار می‌کند. ASR انگلیسی با ۲۲۰۰۰ لغت و ASR فارسی با ۹۰۰۰ لغت کار می کند. ASR از الگوی صوتی کاربر استفاده می‌کند، به این شیوه که در شروع کار، با نمونه‌های صوتی و اطلاعات آماری درباره ساختار زبان وی آموزش[۸] داده می شود.

۳-۲- بخش مدیریت مکالمات (DM)

بخش مدیریت مکالمات (DM) به عنوان قلب سیستم عمل می‌کند و پیام‌ها را میان زیر سیستم‌های مختلف انتقال می‌دهد. خروجی ASR با برچسب هایی نظیر زمان ورود، شماره‌های سریال و سایر اطلاعات مهم دیگر برچسب‌گذاری

---

[۵] Subsystem
[۶] Decode
[۷] Transcript
[۸] Learn

شده و به بخش مدیریت مکالمات فرستاده می‌شوند تا برای نمایش در واسط گرافیکی کاربر[۹] و نیز ترجمه توسط به واحد ترجمه ماشینی[۱۰] انتقال داده شوند. در کنار همه این وظایف، بخش مدیریت مکالمات این اختیار را دارد تا برخی نتایج موجود در لیست خروجی حاصل از پردازش ASR را، بر اساس درجه اطمینان گزارش شده آن‌ها رد کند. این گزینه می‌تواند توسط کاربر تعیین شود.

۳-۳- بخش ترجمه ماشینی (MT)

واحد ترجمه ماشینی یا MT وظیفه ترجمه مکالمات گفته شده بین دو شخص را بر عهده دارد. نکته قابل توجه در این بخش صحت انتقال مفاهیم است.

روش های اولیه ترجمه ماشینی شامل ترجمه آماری عبارات ورودی است، به صورتی که جمله ورودی با استفاده از ترجمه لغت به لغت و سپس تبدیل آن به مدل زبانی مقصد ترجمه می‌شود. این نوع ترجمه، کیفیت مطلوبی فراهم نمی-آورد. از آنجا که صحت تعبیر عبارت ورودی و ترجمه آن مد نظر است، و هم چنین حوزه کاری این سیستم فقط محدود حوزه مسائل پزشکی-درمانی است، بنابراین امکان آن وجود دارد تا بجای ترجمه لغت به لغت، معادل عبارات متداول و مورد استفاده در مسائل پزشکی و ترجمه منتسب به آن‌ها در پایگاه دانش[۱۱] سیستم ذخیره شود. بنابراین در زمان ترجمه، ابتدا میزان نزدیکی مفهوم عبارت ورودی به کلیه عبارات ذخیره شده در پایگاه دانش اندازه‌گیری می‌شود و در صورتی که میزان نزدیکی از حد آستانه بیشتر باشد، عبارت ورودی با عبارت نزدیک به آن در پایگاه دانش، جانشین شود. پس از این مرحله جانشینی، ترجمه ذخیره شده منتسب به آن عبارت، به عنوان ترجمه گفتار ورودی استفاده می‌شود.

در مجموع برای ترجمه ماشینی سیستم جاری، از تکنیک های مختلف ترجمه استفاده شده است که عبارتند از:

۱. کلاسبندی کننده مفاهیم گفتار ورودی بر مبنای ادراک ماشینی
۲. موتور ترجمه به عنوان ماشین مترجم آماری
۳. ترجمه درون زبانی که زیر مجموعه هر دو از انواع قبل است.

---

[۹] Graphical User Interface
[۱۰] Machine Translation Unit
[۱۱] Knowledge Base

بنابراین بخش MT سیستم جاری، در دو حالت[۱۲] کار می کند، که این حالت کاری توسط واحد کلاسبندی[۱۳] تعیین می‌شود. این واحد همانطور که قبلا ذکر شد، در ابتدا سعی می‌کند تا مفهومی را به عبارت ورودی نسبت دهد تا عبارت ورودی را استاندارد کند. برای مثال جمله زیر

"Umm and do you have any headache"

را به مفهوم دیگری نظیر

"Do you have a headache?"

تبدیل می‌کند. اگر این عبارت استاندارد شده در پایگاه دانش واحد کلاسبندی کننده (که حدود ۱۲۰۰ کلاس شامل عبارات است که سیستم با آنها آموزش داده شده است) موجود باشد، ترجمه آن نیز از پایگاه دانش تعبیه شده استخراج می‌شود. اما اگر عبارت ورودی خارج از پایگاه دانش باشد، آن عبارت با استفاده از روش‌های آماری واحد پردازش ماشینی آماری[۱۴] یا SMT ترجمه می‌شود. SMT جملات را به بخش‌هایی می شکند و آنها را جزء به جزء ترجمه می کند. سپس آن ترجمه را سرهم کرده و به عنوان ترجمه عبارت ارائه می‌دهد. این معادل ترجمه لغت به لغت ذکر شده است.

۳-۳-۱- افزایش بازده کلاسبندی

یکی از راه‌های افزایش بازده ترجمه این است که تعداد کلاس های مربوط به مفاهیم را افزایش دهیم تا درصد عبارات شناخته شده توسط کلاسبندی کننده[۱۵] ی سیستم بالا رفته و از صحت ترجمه عبارات اطمینان حاصل شود. اما حتی با اضافه کردن درصد زیادی از عبارات موجود در مکالمه، از وجود واحد ترجمه آماری بی نیاز نخواهیم شود.

برای همین منظور راه‌کارهای افزایش بازده به عبارت زیر خواهند بود:

۱. افزایش دقت کلاسبندی کننده
۲. به دست آوردن مکانیزم موثری برای انتخاب نحوه ترجمه (MT یا SMT)

[۱۲] Mode
[۱۳] Classifier
[۱۴] Statistical Machine Translation
[۱۵] Classifier

۳-۳-۲- کلاسبندی مفاهیم[۱۶] و مدل ادراکی[۱۷]

استفاده از کلاسبندی کننده مفاهیم در ترجمه ماشینی، بر اساس پوشش حوزه مکالمه[۱۸] مورد نظر با تعدادی کلاس از مفاهیم است. برای مثال در حوزه مکالمات پزشکی، مکالمات یک پزشک می تواند به ۱۲۰۰ کلاس از پیش تعریف شده نگاشت شده و مکالمات بیمار می تواند به ۴۰۰ کلاس از پیش تعریف شده نگاشت شود. هر کلاس نماینده‌ای[۱۹] دارد که برای هر گفتگوی ورودی از زبان مبدا (مثلا انگلیسی)، مفهوم مناسب در زبان مقصد (مثلا فارسی) را به دست می‌دهد. در حقیقت کلاسبندی کننده سعی می‌کند گفتگوی ورودی از مبدا را به یکی از مفاهیم مقصد نگاشت کند.

اگر حوزه مکالمه[۲۰] به تعدادی کلاس از مفاهیم نگاشت شود:

$$\aleph = \{C^{(1)}, C^{(2)}, ..., C^{(|\aleph|)}\} \quad (۱)$$

عمل کلاسبندی به صورت ماکزیمم یک تخمین پسین[۲۱] به دست می آید:

$$\hat{C}_t = \arg\max_{C} P(C \mid O_t) \quad (۲)$$

که $C \in \aleph$ است و O_t مشاهده صوتی عبارت گفته شده به زبان مبدا در نوبت t ام و $\hat{C}_t$ مفهوم تخمین زده شده برای آن عبارت است.

در عمل، تخمین بالا در دو مرحله پیاده می‌شود: در مرحله اول، سیگنال صوتی توسط ASR به متن تبدیل می‌شود و سپس یک کلاسبندی کننده متنی، از کلمات موجود در متن حاصل از ASR، به عنوان بردار ویژگی استفاده کرده و متن مزبور را به یک مفهوم نگاشت می‌کند. بدون دانش اولیه در باره مفاهیم، این مساله به یک مساله کلاسبندی بر اساس بیشترین شباهت[۲۲] تقلیل می یابد:

---

[۱۶] Concept Classification
[۱۷] Understanding Model
[۱۸] Dialog
[۱۹] Representative Surface
[۲۰] Dialog Domain
[۲۱] Posteriori Estimation
[۲۲] Maximum Likelihood Classifier

$$\hat{C}_t = \arg\max_C P(\hat{\mathbf{W}}_t \mid C)$$

که در اینجا $\mathbf{W_t}$ برداری از کلمات است که توسط ASR به عنوان متن معادل به دست آمده است. تابع شباهت $P(\hat{\mathbf{W}}_t \mid C)$ می تواند توسط مدل زبانی[۲۳] (LM) ای که منحصرا برای آن زبان ساخته شده، تخمین زده شود. تمامی این مدل های زبانی منحصر شده برای هر مفهوم، مدلی به نام «مدل ادراکی[۲۴]» را می سازند. یک سیستم ترجمه گفتار به گفتار دو طرفه، نیازمند دو کلاسبندی کننده می باشد، مشابه آنچه در (۲) آمده، است. و هر کدام از آن ها باید یک مدل ادراکی از زبان مربوط به خود را داشته باشند. در هر حال، مجموعه مفاهیم با توجه به کاربرد می تواند تغییر کند، به گونه ای که سیستم را اصطلاحا نامتقارن[۲۵] کند.

۳-۳-۳- مدل مکالمه

سیستمی که در بخش قبل توضیح داده شد از اطلاعات مفهومی استفاده نمی‌کند. در یک سیستم دو طرفه نامتقارن عباراتی که فرد کنترل کننده مکالمه (مثل پزشک که جریان مکالمه را به صورت یک مکالمه ی معمول پزشک-بیمار کنترل می کند) بیان می کند، غالبا حاوی اطلاعاتی است که پاسخ دهنده قرار است پاسخ دهد.

یک سیستم واسط دو طرفه را در نظر بگیرید، که هر کدام از طرفین شرکت کننده در مکالمه مجموعه مفاهیم مختص به خود را (نظیر Q-task) داشته باشند. این سیستم نامتقارن، مجموعه مفاهیمی برای بخش کنترل کننده (سمت «الف») خواهد داشت:

$$\Re = \{R^{(1)}, R^{(2)}, ..., R^{(|\Re|)}\}$$

و مجموعه متفاوتی برای قسمت دیگر (سمت «ب»):

$$\aleph = \{C^{(1)}, C^{(2)}, ..., C^{(|\aleph|)}\}$$

[۲۳] Language Model
[۲۴] Understanding Model
[۲۵] Asymmetric

هدف پیدا کردن کلاسبندی کننده جدیدی برای سمت «ب» است، به گونه ای که از تصمیمات اتخاذ شده در سمت «الف» استفاده کند، به این امید که این اطلاعات اضافی دقت کلاسبندی را در سمت «ب» افزایش دهد.

با این فرض که زنجیره وابستگی محدود به یک سیکل گفتگو باشد، کلاسبندی کننده برای قسمت «ب» می تواند به صورت یک تخمین زننده پسین بیشینه[۲۶] مانند شکل زیر باشد:

$$\hat{C}_t = \arg\max_{C} P(C \mid O_t, R_t) \qquad (۳)$$

به صورتی که $C \in \aleph$ و $R_t \in \Re$ و O_t مشاهده صوتی در سمت «ب» در دور t ام صحبت است. به این دلیل که در عمل متن معادل با O_t برای کلاسبندی مورد نیاز است، (۳) را به صورت زیر بازنویسی می کنیم:

$$\max_{C,\mathbf{W}} P(C, \mathbf{W} \mid O_t, R_t) \qquad (۴)$$

$\mathbf{W}$ متن معادل با O_t است. بدون داشتن دانش اولیه R_t، فرمول بالا معادل است با:

$$\max_{C,\mathbf{W}} P(O_t \mid \mathbf{W}, R_t, C) \cdot P(\mathbf{W}) \cdot P(\mathbf{W} \mid C, R_t) \cdot P(C \mid R_t) \qquad (۵)$$

در اینجا بر ای ساده سازی فرمول بالا از دو فرض اولیه استفاده می شود:

۱. فرض اول این است که

$$P(O_t \mid \mathbf{W}, R_t, C) \approx P(O_t \mid \mathbf{W})$$

به این معنا که مشاهده صوتی به مفهوم انتخابی از هر دو طرف مکالمه، ارتباطی ندارد.

۲. همچنین فرض می شود که متن معادل با گفتار سمت «ب» و مفهوم سمت «الف» مستقل از هم باشند، یعنی:

$$P(\mathbf{W} \mid C, R_t) = P(\mathbf{W} \mid C)$$

[۲۶] Maximum Estimator Posterior

این فرض ها باعث می‌شود تا بتوان معادله (۵) را به دو مرحله ی بیشینه سازی[۲۷] تقسیم کرد:

$$\max_{C,\mathbf{W}} P(O_t \mid \mathbf{W}, R_t, C) \cdot P(\mathbf{W}) \cdot P(\mathbf{W} \mid C, R_t) \cdot P(C \mid R_t) \quad (۶)$$

$$\hat{C}_t = \arg\max_{C} P(\hat{\mathbf{W}}_t \mid C) \cdot P(C \mid R_t) \quad (۷)$$

$\hat{w}_t$ در (۶) می تواند خروجی ASR باشد، که این ASR دارای مدل های صوتی و زبانی‌ای است که می توانند که به ترتیب $P(O_t \mid W)$ و $P(W)$ را تخمین بزنند. معادله (۷) نشان می دهد که مفهوم عبارت ذکر شده در سمت «ب» می-تواند توسط خروجی ASR سمت «الف» و مفهومی که برای آن انتخاب شده است، تخمین زده شود. با این که تخمین زننده ی (۲) تنها از اطلاعات یک سمت استفاده می‌کند، (۷) از ارتباطات و وابستگی‌های آماری مفاهیم هر دو سمت مکالمه استفاده می کند، یعنی $P(C \mid R_t)$ می توان این وابستگی را توسط مدل مکالمه تخمین زد.

در عمل معادله (۷) به صورت زیر عمل تخمین مفاهیم را انجام می دهد:

$$\hat{C}_t = \arg\max_{C} P_U(\hat{\mathbf{W}}_t \mid C) \cdot [P_D(C \mid R_t)]^{\gamma} \quad (۸)$$

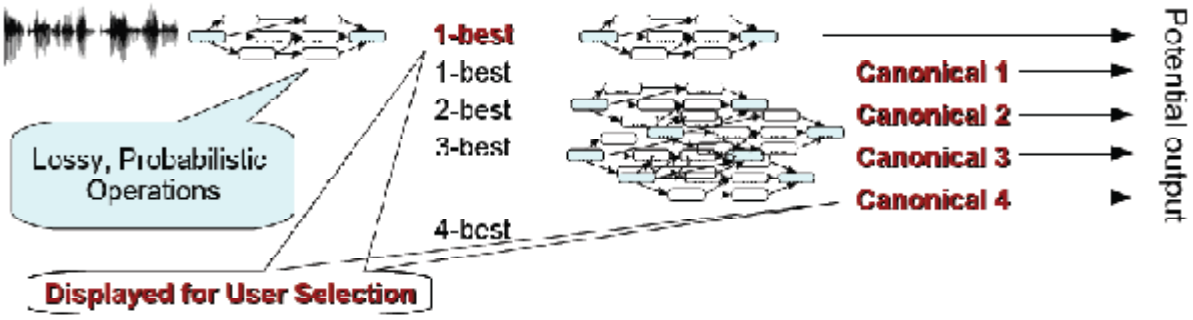
که P_U و P_D به ترتیب مدل ادراکی و مدل مکالمه هستند. این توابع تخمین ناقصی از معادله (۷) هستند. ضریب توانی γ برای تقویت یا تضعیف اثر مدل مکالمه آورده شده است.

۳-۴- واسط گرافیکی کاربر (GUI)

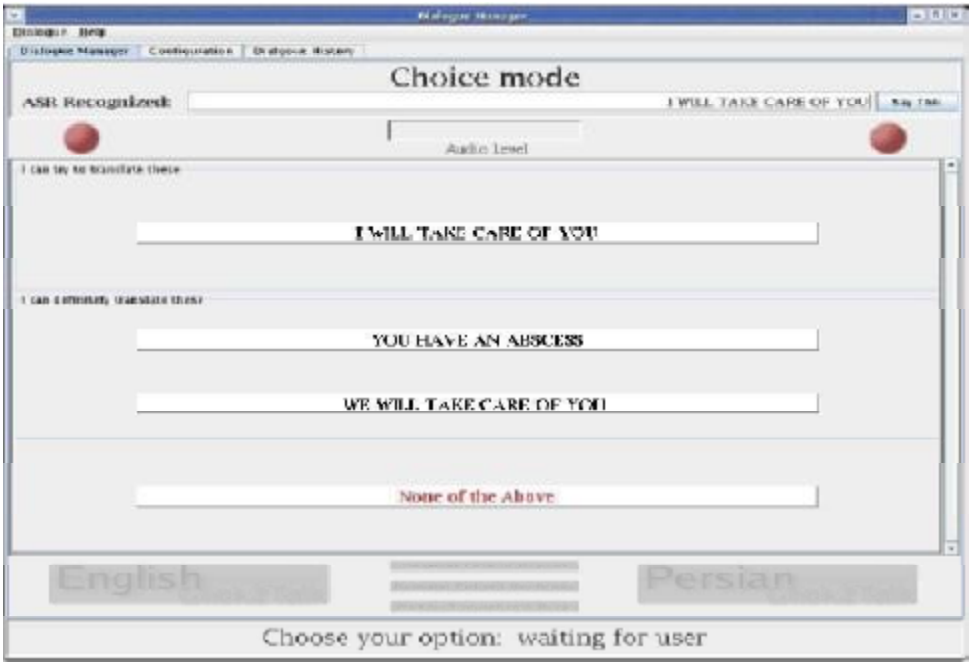
تا کنون تمامی کامپوننت های ذکر شده برای کاربر نامحسوسند، به جز واسط گرافیکی (که شامل راهنمای کاربر[۲۸] و راهنمای سیستم[۲۹] می باشد). از GUI برای بالا بردن دقت ترجمه ماشینی استفاده می‌شود. به این صورت که پس از این که بیمار صحبت خود را تمام کرد، آن را پردازش کرده و سپس از پزشک خواسته می شود تا از بین مفاهیم به دست آمده برای ترجمه گفتار وی، بهترین و مرتبط ترین مفهوم را انتخاب کند.

[۲۷] Maximization
[۲۸] User Manual
[۲۹] System Help

این سیستم به این شیوه کار می‌کند که برای مثال، ۵ گزینه از ترجمه های متنوع عبارت ورودی بیمار به پزشک نشان داده می شود. ترجمه اول با استفاده از ترجمه آماری صورت می گیرد و ۴ ترجمه دیگر با استفاده از مدل رقابتی واحد کلاسبندی به دست می‌آید. این ۵ گزینه در اختیار پزشک قرار می گیرد و وی با استفاده از دانش اولیه و سابقه ی گفتگو بهترین ترجمه را انتخاب می کند (شکل ۲). شکل ۳ شمای کلی نرم افزار را در زمان انتخاب بهترین گزینه ترجمه نشان می دهد. شکل ۴ تاریخچه گفتگو را نشان می دهد.



شکل ۲ یکی از ترجمه ها بر اساس ترجمه آماری و ۴ ترجمه دیگر بر اساس مدل کلاسبندی انجام می شود



شکل ۳ نحوه انتخاب بهترین ترجمه

شکل ۴ تاریخچه گفتگو

۳-۵- بخش سنتز متن به گفتار (TTS)

واحد سنتز متن به گفتار[۳۰] یا TTS متن ترجمه شده را به صورت آوایی زبان مقصد تولید می کند. TTS 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 می‌کند و سلسله ای از گفتارهای از پیش ضبط شده برای آن کلمات پخش می‌شوند.

۳. برای کلماتی که صورت از پیش ذخیره شده ای برای آن ها وجود ندارد، یعنی لغات غیر معمول، سنتز در سطح Diphone انجام می‌شود.

[۳۰] Text To Speech synthesizer

۴- معماری سیستم

در معماری سیستم، تمامی اجزاء با استفاده از شبکه های کامپیوتری به یکدیگر قابل اتصال هستند، بنابراین می-توانند در ماشی‌ن‌های متفاوت و مجزا اجرا شوند. بعلاوه اینکه همگی این اجزاء قابل اجرا بر روی سیستم عامل Windows و Linux هستند.

برخی از کامپوننت‌های موجود از چندین ریزکامپوننت ساخته شده اند، برای مثال کامپوننت TTS متناسب با روش سنتز متن ورودی، از واحدهای سنتز در سطح عبارت، سطح کلمه و سطح Diphone ساخته شده است که هر کدام از این کامپوننت ها نیازمند پردازش جداگانه است.

در معماری سیستم جاری، همه پیام ها با برچسبی که شامل اطلاعات مربوط به میزان داده، مبدا و مقصد پیام است برچسب می خورند و به تمامی کامپوننت ها ارسال[۳۱] می شوند (شکل ۵). این موضوع عمل یکپارچه و نیز مونیتورینگ کانال‌های ارتباطی داخلی را برای DM میسر می‌کند؛ به صورتی که می‌تواند کل سیستم را برای درخواست اطلاعات کاربر متوقف کند. بیشترین نوع درخواست های سیستم از کاربر شامل تکرار و تایید عبارت و رفع ابهام آن با استفاده از انتخاب بهترین عبارت از لیست گزینه ها است که پاسخ کاربر می‌تواند از طریق صوت و یا واسط گرافیکی به سیستم رسانده شود.

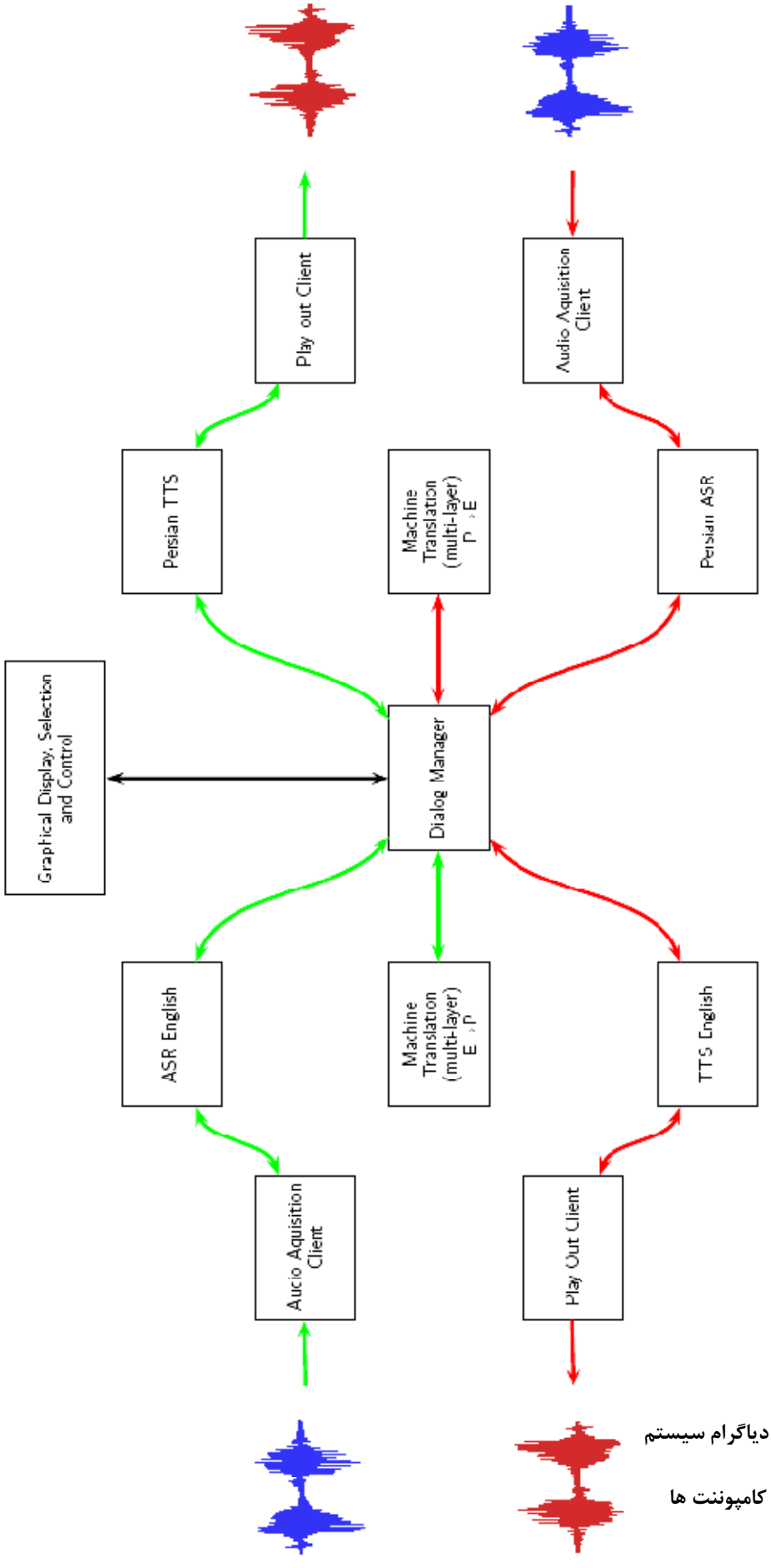
برای ASR انگلیسی از پردازنده Sonic دانشگاه Colorado استفاده شده است که از ابتدا توسط داده های LM جمع‌آوری شده از منابع مختلف -نظیر ذخیره دیالوگ های پزشک-بیمار جمع آوری شده از دانشجویان پزشکی- آموزش دیده است. برای مدل صوتی فارسی به دلیل کمبود داده‌های صوتی برچسب خورده در این زمینه، از مدل صوتی انگلیسی با استفاده از یک نگاشت زیرآوایی[۳۲] بین آواهای دو زبان، انجام گرفته است. هم چنین از مجموعه اصوات فارسی به نام FARSDAT و نیز اصوات گفتاری افراد فارسی زبان استفاده شده است.

[۳۱] Broadcast
[۳۲] Sub-phonic

مسیری که برای گفتار فارسی و انگلیسی پیموده می‌شود تا گفتار بیمار فارسی زبان به گوش پزشک انگلیسی زبان برسد، و برعکس در شکل ۶ آمده است.

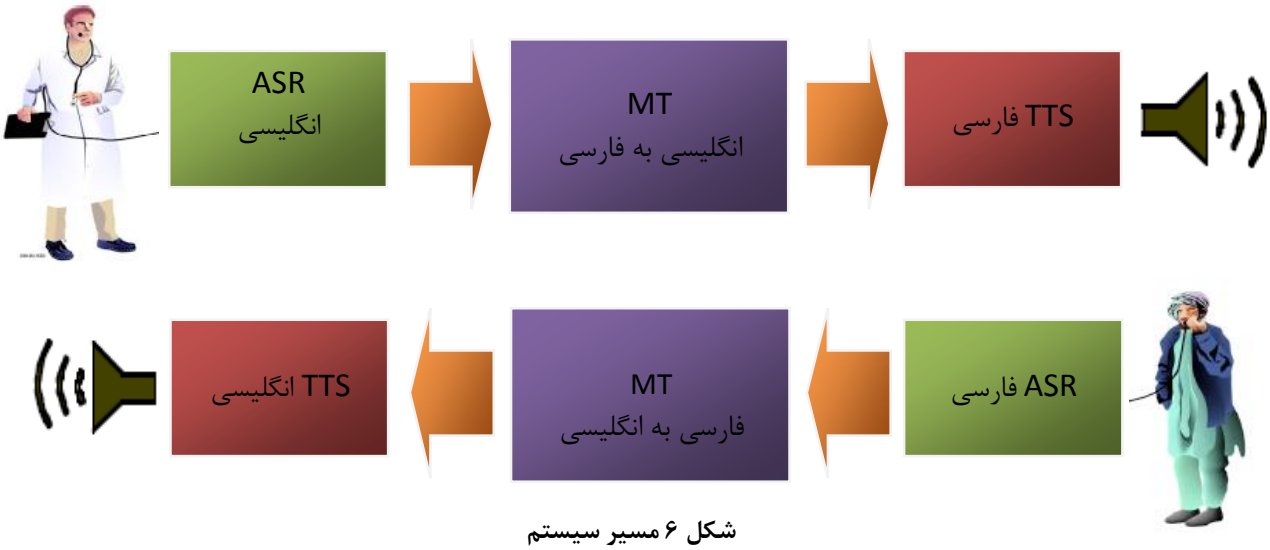
بخش MT، همانطور که در قبل گفته شد، از کلاسبندی کننده و ماشین ترجمه آماری تشکیل شده است. کلاسبندی کننده انگلیسی تقریبا از ۱۴۰۰ کلاس متفاوت تشکیل شده است که هر این کلاس ها حاوی عبارات متداول و استانداردی است که در محاوره ها پزشک- بیمار استفاده می شود. هر کلاس دارای مجموعه بسیار بزرگی از تفاسیر معادل عبارات داخل کلاس نیز می باشد، به صورتی که اگر پزشک یکی از آن تفاسیر را به کار ببرد، سیستم عبارت معادل آن را تشخیص دهد و آن را با استفاده از جدول مرجع[۳۳] به فارسی ترجمه کند. اگر کلاسبندی کننده نتوانست معادلی را برای عبارت گفته شده پیدا کند، به گونه‌ای که حد آستانه اطمینان را برآورده سازد، از MT آماری یا SMT استفاده می شود. SMT از لیست n تایی از عبارات معادل بین دو زبان مبدا و مقصد استفاده می کند. همانند ASR بازده این بخش ارتباط تنگاتنگی با تعداد داده های آموزش دارد.

[۳۳] Lookup Table



شکل ۵ بلوک دیاگرام سیستم

و نحوه اتصال کامپوننت ها



شکل ۶ مسیر سیستم

۵- جمع آوری داده

برای جمع آوری داده در زمینه اصوات فارسی منابع بسیار کمی در دسترس بوده است. تنها پایگاه داده صوتی به زبان فارسی، پایگاه FARSDAT است. این پایگاه داده متشکل از ۲۰ جمله است که توسط ۳۰۰ نفر با سن، جنسیت، سطح تحصیلات و لهجه متفاوت به زبان فارسی خوانده شده است، که در مجموع شامل ۶۰۰۰ عبارت را شامل می‌شود. این تعداد برای کاربرد تشخیص صوت ناکافی است، بنابراین برای به دست آوردن پایگاه داده غنی‌تر، از ۳۰۰ عبارت استاندارد که در مکالمات پزشکی به کار برده می‌شود، استفاده شده است. ای عبارات توسط تعداد زیادی از افراد فارسی زبان خوانده و ذخیره شده است.

بدین ترتیب پایگاه داده ای بالغ بر ۳۰۰ هزار کلمه زبان فارسی و انگلیسی به دست آمده است.

۶- نتایج

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

۱. ۱۵۰ میلیون کلمه مجزا که از Web به دست آمده است و برای حوزه کاری سیستم فیلتر نشده است

۲. ۱۵۰ میلیون کلمه مجزا از Web که برای حوزه کاری سیستم فیلتر شده است

۳. پایگاه داده PPL

۴. پایگاه داده BLEU

۵. پایگاه داده LPU

نتایج خطای تشخیص برای هر کدام از پایگاه های داده فوق در جدول ۱ آمده است

جدول ۱ خطای تشخیص در سیستم Transonic

	۱۰K	۲۰K	۳۰K	۴۰K
No Web	۱۹.۸	۱۸.۹	۱۸.۳	۱۷.۹
All Web	۱۹.۵	۱۹.۱	۱۸.۷	۱۷.۹
PPL	۱۹.۲	۱۸.۸	۱۸.۵	۱۷.۹
BLEU	۱۹.۳	۱۸.۸	۱۸.۵	۱۷.۹
LPU	۱۹.۲	۱۸.۸	۱۸.۵	۱۷.۸
Proposed	۱۸.۳	۱۸.۲	۱۸.۲	۱۷.۳

نتایج نشان می دهد که سیستم برای داده های حاصل از وب که برای این حوزه کاری فیلتر نشده اند، سیستم 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 و به گفته توسعه دهندگان این سیستم، نتیجه فاجعه بار (تشخیص نادرست و زیانبار بیماری) توسط این سیستم گزارش نشده است.

۷- مراجع

[۱]. Emil Ettelaie, *"Cross-lingual Dialog Model for Speech to Speech Translation", ICSLP, INTERSPEECH* ۲۰۰۶

[۲]. *Shrikanth Narayanan, "Speech Recognition Engineering Issue In Speech To Speech Translation System Design For Low Resource Languages And Domains*", IEEE, ۲۰۰۶

[۳]. JongHo Shin, "*User Modeling in a Speech Translation Driven Mediated Interaction Setting*", Viterbi School of Engineering, ۲۰۰۶

[۴]. Emil Ettelaie, "*Transonics: A Practical Speech-to-Speech Translator for English-Farsi Medical Dialogues*", Proceedings of the ACL Interactive Poster and Demonstration Sessions, June ۲۰۰۵

[۵]. S. Narayanan, *"The Transonic Spoken Dialogue Translator: An Aid For English-Persian Doctor-Patient Overviews"*, University of Southern California, ۲۰۰۴

[۶]. Panayiotis G. Georgiou, *"An English-Persian Automatic Speech Translator: Recent Developments In Domain Portability And User Modeling"*, University of Southern California

[۷]. http://sail.usc.edu/transonics/s۲s.php